Beyond the Boundary

An analysis of verisimilitude and causal ontology of scientific claims

Ætiologies of developmental dyslexia as a case in point

Fred Hasselman

May 18, 2015

The author conducted the research and wrote the dissertation while he was affiliated to:
School of Pedagogical and Educational Science,
School of Psychology and Artificial Intelligence,
Behavioral Science Institute: Learning & Plasticity

Links to **pdf** files of each chapter and the **data** and code to reproduce the **analyses** and **figures** presented in each chapter, are avaliable through a public project hosted at the COS - Open Science Framework. The links point to version controlled files hosted on figshare and GitHub.

https://osf.io/rupg8/

http://figshare.com/authors/Fred_Hasselman/520930

https://github.com/FredHasselman

Printed by PrintPartners Ipsamp b.v., Enschede, The Netherlands

Beyond the Boundary

An analysis of verisimilitude and causal ontology of scientific claims

Ætiologies of developmental dyslexia as a case in point

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen op vrijdag 22 mei 2015
om 10.30 uur precies

door

Frederik Willem Hasselman
geboren op 25 december 1973
te Geldrop

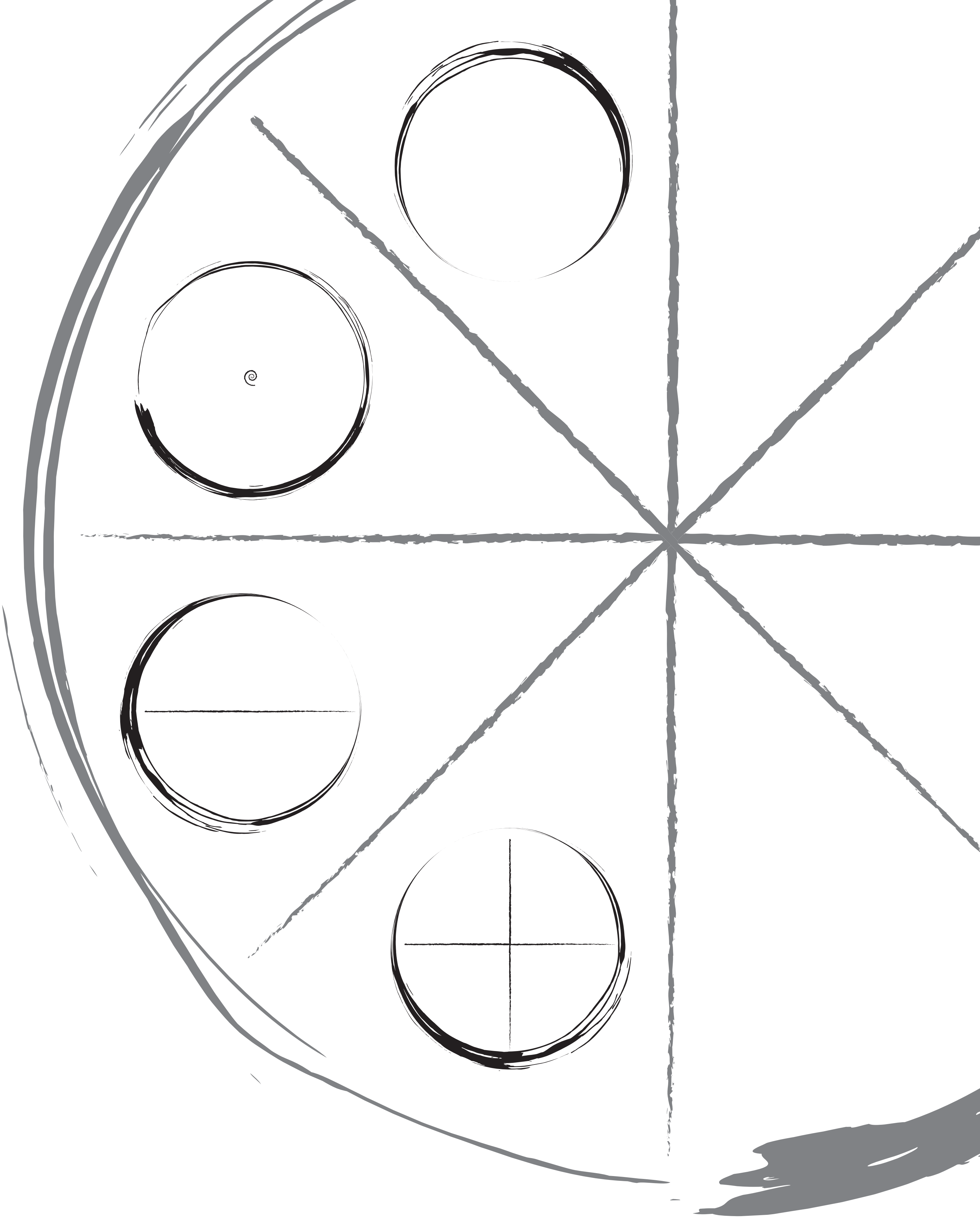
Promotoren:

Prof. dr. A.M.T. Bosman
Prof. dr. L.T.W. Verhoeven

Manuscriptcommissie:

Prof. dr. M.V.P. Slors
Prof. dr. D.A.V. van der Leij (Universiteit van Amsterdam)
Dr. J.G. Holden (University of Cincinnati, Verenigde Staten)

Contents



"A beginning is the time for taking the most
delicate care that the balances are correct."

- Fom *Manual of Muad'Dib* by Prinses Irulan (Herbert, 1965)

The Pagan elements are 4 + 1. Though many neo-paganists use the Hellenic or Alchemical triangular symbols for the Earth, Air, Water and Fire, the 5th element is almost always the circle divided into 8 segments, signifying the element of Spirit, the Unifying force. Sometimes it is also called Heaven or ***Æther*** after Aristotle's **quintessence**, though Plato associates it with the world of Forms (Ideai). The five elements may be aligned along points of a pentagram, a geometrical figure which plays an important role in many Pagan rituals. This practice is thought to have originated with the early Pythagoreans (de Vogel, 1966) although annotated pentagrams have been found on coins and artifacts as early as the 4th century BCE. It must be noted that the Chinese 5 elements are also often diplayed along the points of a pentagram.

Apocrypha

Herbert, F. (1965). Dune. Chilton Book Company, Philelphia, USA.

de Vogel, C. J. (1966). Pythagoras and Early Pythagoreanism. Van Gorcum, Assen, The Netherlands.

Notes

Preface

This book is not about developmental dyslexia.

This book is about a science.

A science in perpetual crisis.

The hardest science.

The science that boldly sets out to accumulate knowledge about Human *Nature*.

Therefore, this book is about a *Natural* Science.

The majority of the empirical and theoretical content of this book may be summarised as an attempt to evaluate knowledge accumulated about a well studied phenomenon of human nature (developmental dyslexia), from a perspective on theory evaluation that has been very successful in advancing scientific knowledge about natural phenomena. It comes down to this: Empirical inquiries, inferences, and conjectures should be guided by theories of principles that can be truth-like or false, not by theories that can be constructed such that virtually any outcome constitutes a corroboration of its `predictions'.

The purpose of this preface is to point to a crisis in the science of human nature, that is much more fundamental than a crisis of confidence in the validity of statistical inferences or empirical record due to questionable research practices.

Science is one of the arts that studies the structure of reality. The difference between the scientist and the artist is that the latter is free to choose the methods that constrain the way reality may be observed and expressed. The scientist has just one option:*the scientific method*. The scientific method is not a tool for building an industry that produces facts about reality. It is a tool designed to prove explanatory claims about the structure of reality wrong. Science should not produce anything other than *more questions*.

Sane as it Ever Was

The Historical Meaning of the Crisis in Psychology

A Tapestry Wanted

In an article entitled "Has Psychology Failed?" Joseph Jastrow reflects on whether prominent scholars such as William James and Stanley Hall had been right to be pessimistic about the advancement of psychology since the days of its conception as an empirical science. He notes the following:

> ``The present text-book chaos is the work of drifting pilots. They leave the student with the impression of a patch-work quilt whereas actually the mind is a tapestry.'' (Jastrow, 1935, p. 268)

In this preface to the chapters of a book that concern the specific case of evaluating competing ætiologies of developmental dyslexia, I conjecture that the general failure to advance the empirical social sciences as a natural science lies in the not so recent past. The current crisis of confidence in the empirical record of the social and life sciences is not due to the (ab)use of inferential statistics as a means to haul in huge quantities of scientific knowledge, but is more likely due to the inability of the scientific community to stop expanding the patch-work quilt by suggesting a new theory (or paradigm) for each freshly caught fact. That is, this crisis concerns the curious case of an empirical science in which the very thing that separates this branch from the other disciplines –knowledge inference by means empirical tests of theoretical claims– seems to have little or no authority at the level of evaluating the truth-likeness (verisimilitude) of theories.

The conjecture: If empirical disciplines of social science experienced crises of confidence before statistical hypothesis testing was adopted as the main tool for scientific inference, it is unlikely that inferential statistics are the true cause of the current crisis. Reforms exclusively aimed at improving this type of scientific inference, although necessary, are unlikely to be successful in preventing another crisis from happening again in the next decade. Based on an examination of historical commentaries on the state of theory evaluation in psychological science, I will suggest that the only way forward for the empirical social sciences in general, is to join the tapestry weaving guild and establish a consensus science, a science in which "theory and data speak more for themselves" (Fanelli, 2010). In the 'softer' fields of empirical science theoretical disputes mainly concern circular 'special sauce' arguments: "You did not evidence the phenomenon I study, because you do not know what it takes to evidence the phenomenon I study and neither do I when you ask me to be explicit about it a-priori".

This transition will not happen overnight, but as I will argue in what follows, the claim of being young and still maturing as a science as an excuse for its delayed occurrence is untenable. The importance of transforming into a consensus empirical science was recognised in the earliest Principia of Psychology. Moreover, I believe it is essential to start this discussion right now, because the Open Science movement has all the potential to play a crucial role in laying down the foundations for a discipline that studies human nature as a natural science within the next decade.

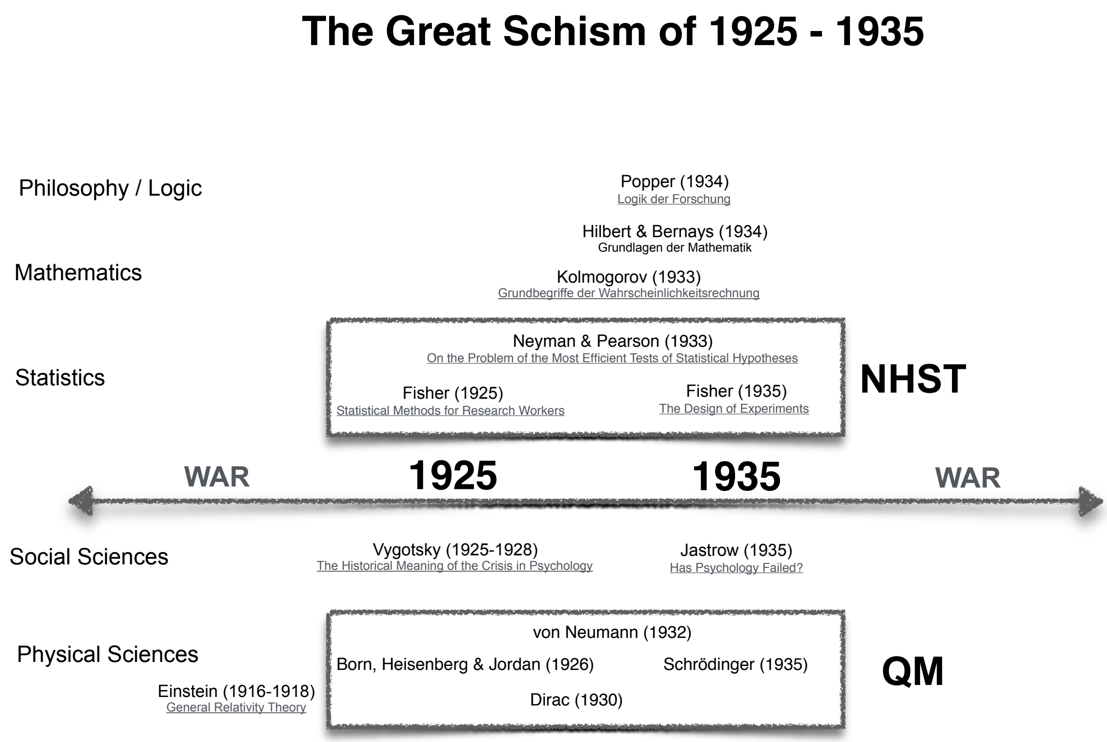


Figure 0.1 – Intriguing times in the history of science. The great schism between the harder and softer disciplines.

The Reports of your Immaculate Conceptions are Greatly Exaggerated

The other day, when I released my conjecture into the wild in front of the coffee machine, a senior colleague objected: "What do you mean, the crises in psychology before the current one? This is the first crisis I have ever experienced!" Well, to give a hint, the subtitle "The Historical Meaning of the Crisis in Psychology" is not my invention[1]. It is the title of a 1925 manuscript attributed to Lev Vygotsky (for details see Zavershneva, 2012) and is one of many analyses of the 'psychological crisis' published in the pre-NHST era, among which are works by Karl Bühler, Kurt Koffka, William Stern, Mary Whiton Calkins, Edmund Husserl and Kurt Lewin (cf. Yasnitsky, 2011). Upon hearing those names, one might object those works could not possibly concern anything related to the crisis in 'modern' psychology.

For convenience I will mark the years 1925-1935 as a transitional period after which the current statistical era began (see Figure **0.1**), it includes the publication of the famous works by Fisher (1925, 1935) and Neyman and Pearson (1933). Before I discuss the nature of the pre-NHST crisis, its relation to the current crisis, and the unique role the open science movement could play in its resolution, I need to put a date on the birth of psychological science as a modern empirical discipline:

> ``Modern Psychology surely began, not ``three or four years ago,'' with the publication of the Willenshandlung, –but some forty years ago, with Fechner's notion of the definite functional correlation of psychical with physical processes. The modern psychologist is the experimental psychologist.'' (Titchener, 1893, p. 456, emphasis added)

[1]Neither is the main title, see http://www.comicvine.com/cerebus-70-sane-as-it-ever-was/4000-124733/

Box 0.1: A (non-exhaustive) comparison between announced origins of modern psychology (hyperlinks to text in electronic document) and origins of modern versions of other disciplines of science.

Herbart (1824; 1825)	Psychology as a Science (Synthetic & Analytic Volume)	**Modern life science / medicine** (1855) Remak / Virchow discover the building block of life, cause of pathology: The cell
Combe (1834)	Principles of Physiology	
Spencer (1855)	The Principles of Psychology	**Modern Biology:** (1859) Darwin's famous book was published
Fechner (1860)	The Elements of Psychophysics	Spencer: "survival of the fittest" (1864)
Carpenter (1874)	Principles of Mental Physiology	
Münsterberg (1888)	Die Willenshandlung	**Modern (Theoretical) Physics:** (1865)
James (1890)	The Principles of Psychology (Volumes I & II)	Maxwell's dynamical theory of the electromagnetic field
Wundt (1874; 1897)	Principles of Physiological Psychology Outlines of Psychology	**Modern (Cognitive) Neuroscience:** (1868) F.C. Donders' mental chronometry
Titchener (1898)	The Postulates of a Structural Psychology	**Modern Mathematics:** (1859-1900) Riemann, Minkowski, Poincaré, Hilbert

This suggests experimental psychological science is of the same generation as many other modern disciplines of science, some of which can be considered as belonging to the 'harder' disciplines. Biology, physiology and medical science all began in their modern form in the middle of the 19th century when the cell was discovered as a fundamental building block of living organisms and a cause of pathology (cf. Remak, 1855; Virchow, 1855). Naturally, Darwin's famous book published in 1859 set the stage for a whole range of new areas of scientific inquiry (see Figure **0.1**).

The conclusion must be that the systematic study of psychological phenomena using the scientific method Novalis described in 1798 as:

> ``Hypotheses are nets; only he who casts them will catch. Was not the discovery of America the result of a hypothesis? Long, and above all, live the hypothesis - Only she will stay forever new, no matter how often she defeats herself.''

The casting of nets to catch some facts, has been practiced for at least 160 years, with or without the help of statistical inference. Herbert Spencer published one of the first "The Principles of Psychology" and the opening chapter, "A Datum Wanted", stresses the importance of achieving consensus about the object of study in order to establish a genuine science of psychology. Spencer also coined the term 'survival of the fittest' in the "Principles of Biology" (1864). Therefore, measured in chronologically ordered units of Principia produced by Herbert Spencer (of which there are at least 6, covering psychology, sociology, biology, education, ethics and philosophy of science), modern psychological science is older than modern biological science.

2 (±7) Some Limits on the Capacity to Produce Principles, Please!

Spencer's "The Principles of Psychology" triggered exactly the opposite response of what he had wished for. A number of rather different Principia for psychology were published in the second half of the 19th century, all of them were intended to lay the foundations for a study of psychological phenomena as a natural science, but a consensus datum or theoretical framework was never found.

Arguably the most famous of the Principia is William James' "The Principles of Psychology, Vol. I & II" (1890a; 1890b). Other Principia, were Combe's "Principles of Physiology" (1834, and later editions), Fechner's "The Elements of Psychophysics" (1860/1912), Carpenter's "Principles of Mental Physiology" (1874), Münsterberg's "Die Willenshandlung" (1888), then there is of course the "Outlines of Psychology" (1896/1897) and "Principles of Physiological Psychology" (1874/1904) by Wundt and "The Postulates of a Structural Psychology" (1898) by Titchener. Other foundational texts printed after the turn of the century could be added. The behaviourists certainly produced the most formal theories of behaviour: Watson, Thurstone, Hull, Thorndike, Pavlov, Lashley. Other scholars from the pre-NHST era known for initiating (or terminating) foundation debates are Cattell, Baldwin, Dewey, Hall, Allport, Binet, Galton, Pearson, Mead, Lange, Koffka, Freud and Jung.

My guess is few scholars will have studied more than 2 of the primary Principia over the course of their career. I immediately confess my number came up negative about four years ago (when I discovered some Principia I did not know even existed) and it has been very difficult to improve on that number. So before embarking on a reading spree of classical texts (most of which are freely available on-line), let me suggest another strategy. A good place to start is by reading the commentaries, or reviews of these works, of which there are many. For example, here is what George Trumbull Ladd had to say about William James' Principia:

> ``Of the conception of psychology, its nature, problems, and method, which is proposed in these volumes, and of the defence in detail of this conception, the following statements seem to me true: The conception is such, and so narrow, that a consistent adherence to it compels us to admit the utter impossibility of establishing psychology as a natural science. It excludes almost all the really scientific data and conclusions; it includes only those data and conjectures which are most remote from genuine science.'' (Ladd, 1892, p. 28, emphasis added)

Ladd finds himself in good company when he concluded it would be impossible to establish psychology as a natural science based on William James' work. Here is what the author himself had to say about his two volume opus magnum:

> ``No one could be more disgusted than I at the book. No subject is worth being treated in more than 1000 pages. Had I ten years more, I could rewrite it in 500; but as it stands, it is this or nothing –a loathsome, distended, tumefied, bloated, dropsical mass, testifying to nothing but two facts: 1st, that there is no such thing as a science of psychology, and 2nd, that W.J. is an incapable.'' (James, 1920
> 2008, pp. 293–294, emphasis in original)

That's a quote you don't see in textbooks very often. It does reveal how serious the pre-NHST crisis in psychology actually was. A view emerges of a discipline in search of first principles that is clueless about how to decide whether they have actually been found.

Blinded by Downward Seepage

The most comprehensive analysis of the state of pre-NHST psychological science must be Lev Vygotsky's 'crisis in psychology' manuscript. The manuscript was never published in his lifetime (much later in collected works, e.g., Rieber & Wollock, 1997), but improved and amended versions of his analysis and conjectures appeared in the second half of the 1920s, as parts of articles on other subjects (cf. Zavershneva, 2012). The manuscript is more than just an analysis, it provides directions for a future psychology that makes sense today, for instance the interventionist methodology (Sannino & Sutter, 2011). According to his student Luria, Vygotsky had basically studied all the Russian and Western Principia of psychology that were available at the time (cf. Yasnitsky, 2011), the entire patch-work quilt.

One of the causes of the crisis suggested by Vygotsky is intriguing, especially in the context of the recent contribution to the Open Science Collaboration blog by Denny Borsboom (Borsboom, 2013) on Theoretical Amnesia:

> ``When one mixes up the epistemological problem with the ontological one by introducing into psychology not the whole argumentation but its final results, this leads to the distortion of both.'' (Vygotsky, 1925). Roughly translated, even before the rules of statistical inference governed the accrual of scientific knowledge (epistemology), some form of theoretical

> amnesia was identified that was caused by an exclusive focus on observable effects ("final results"), without investigating the derivational chain of theoretical propositions ("whole argumentation") that actually led to the observation of those empirical phenomena.

This situation resembles a reverse "Ramsified upward seepage" (Meehl, 1990). Paul Meehl used 'upward seepage' to explain how logicians account for the implicit empirical, or, real-world content that formal, abstract theoretical propositions acquire (the answer is: by means of a Ramsey sentence). Remember that in scientific theorising there is just a one-way logical route from a derivation based on formal propositions to an observational constraint whose veracity can be empirically appraised, not the other way around. In modern psychological science there appears to be a blinding by 'downward seepage', in which empirical phenomena acquire theoretical content. Like theoretical amnesia, a scientist who is blinded by downward seepage, is generally unaware that EVERY empirical inquiry in science is guided by a derivation chain of theoretical propositions. Discovery of unexpected phenomena is of course possible when a scientist pokes around in the specific domain of reality that presents itself due to his or her theoretical tunnel vision. Novalis was right to observe that the inquiries that led to the discovery of America (that is, by Columbus), were based on propositions, hypotheses about the constituents of reality and their properties: "It is the theory which decides what we can observe" (Einstein as quoted by Heisenberg, 1971, pp. 62–63).

The consequence of blinding by DS is that theoretical constructs escape falsification by repeated application of the scientific method, simply because they are no longer perceived as a theoretical proposition whose truth-likeness (verisimilitude) should be questioned. This resembles a fallacy described by William James:

> ``The great snare of the psychologist is the confusion of his own standpoint with that of the mental fact about which he is making his report. I shall hereafter call this the `psychologist's fallacy' par excellence.'' (James, 1890a, p. 196)

The difference is that instead of mental facts or experiences, blinding by DS concerns more abstract theoretical constructs. Some examples: Phoneme representation, working memory, mental lexicon, cognitive load, attention, motor program, predisposition, competence, information processing, true effect, and the most important one, event probability. These are all theoretical propositions used to describe constituents of reality and the laws that connect them in a theoretical framework, formalism, or philosophy. A curious thing happens when you question their existence in front of empirical scientists. In many cases you will get a response along the lines of: "Really? Try to explain [insert complex human behaviour] without using [insert theoretical proposition that invaded reality]". Do try this at home, in the classroom, or in front of the coffee machine!

Why Psychology is the Hardest Science: The Thin Ontic Line

To many cognitive scientists it will be inconceivable to explain something like chess without assuming mental representations and a database metaphor of human memory as actual constituents of reality. However, to date, there is no evidence to suggest mental representations should be awarded an ontological truth status (e.g., Haselager, de Groot, & van Rappard, 2003) and efforts to find memories as discrete states of the central nervous systems haven't advanced much since Lashley announced his failure to find the engram (e.g., Dudai, 2004). It's not the case the existence of these constructs has been shown to be false, their existence is often assumed as science fact and this is rarely rigorously tested experimentally. As for the mental representation, there isn't even a formal definition available that can be used to provide an empirical existence proof. That reeks of classical metaphysics, not empirical science! The fact that a practitioner of science might find it inconceivable to dispense with a theoretical construct as a constituent of reality, should be irrelevant for guidance of attention of empirical inquiries or theory evaluation. The verisimilitude of a construct can only be evidenced by an immaculate track record of corroborative events obtained by rigorous testing of all the theoretical propositions in the derivational chain, nothing more, nothing less.

Psychological science is the hardest science, not because human behaviour is more context sensitive, or more retro- or prospective than any other animal that can pass the mirror test, even less so because one just can't play chess with turbulence and explaining spatiotemporal chaos is child's play compared to explaining chess. All those complexities can be addressed by a science of psychology seeking scientific explanation based on first principles and universal law and if you do not believe that is possible we should stop calling it a science right now. The reason it is hard to transform into a consensus science that studies human nature is due to the thin ontic line that separates phenomena that should be predicted by a formal scientific theory, from the phenomenological experience of reality as it unfolds in front of our mind's eye. The very first sentences of the very first chapter, of the very first principles of modern psychology already provide a complete description of this problem:

> ``§ 1. The postulates and axioms prefacing our expositions of exact science—our works on Geometry and our Mechanical Treatises—are received on the direct warrant of consciousness that they are indisputable. Similarly with all that we regard as objective truths; whether known immediately by simple intuitions, or mediately by the series of intuitions constituting a deductive argument. But when from objective truths we pass to subjective ones—when from the outer phenomena cognized, we turn to the inner phenomena presented by the act of cognition—when, after analysing knowledge, we begin to analyse that which knows, we are met by the question—What is here our test of validity? Consciousness vouches for the truth of propositions concerning external relations; but what shall vouch for the truth of propositions concerning those internal relations which constitute the phenomena of consciousness?'' (Spencer, 1855, p. 8)

The primary Principia each take position with respect to where this line between phenomena of the mind and phenomena of the physical world should be drawn. This demarkation also implicates whether the mind and subjective experience should be considered a valid object of scientific inquiry. The pre-NHST crisis in empirical psychology concerns the inability to achieve consensus on these matters and is therefore philosophical in nature. It is a crisis that could not be resolved by a laboratory experiment or field study, but requires formal theorising. This insight is indeed described by Vygotsky in the new direction he envisioned for psychological science:

> ``Such a system has not yet been created. We can say with confidence that it will not arise out of the ruins of empirical psychology or in the laboratories of reflexologists. It will come as a broad biosocial synthesis of the theory of animal behavior and societal man. This new psychology will be a branch of general biology and at the same time the basis of all sociological sciences. It will be the knot that ties the science of nature and the science of man together. It will therefore, indeed, be most intimately connected with philosophy, but with a strictly scientific philosophy which represents the combined theory of scientific knowledge and not with the speculative philosophy that preceded scientific generalizations.'' (Vygotsky,1925/1997, p. 61).

Oh no! Empirical psychology was in ruins back then as well. Could it be that the speculative philosophy that preceded scientific generalisations was replaced with speculative statistical generalisations in order to revive empirical psychology from its ruins, without connecting to any strict philosophy of science? We should have listened to Vygotsky. the causes of not being able to decide between the veracity of theoretical claims he described in 1925 do not appear to be very different from the current crisis: The rules of inference, whether based on some arcane metaphysical philosophy, or statistics, do not suffice to advance the scientific knowledge base of an empirical science that studies psychological phenomena.

To summarise so far, the pre-NHST crisis in psychology shares with the current one the inability of the scientific community to produce a formal, consensus definition of what it actually studies and more importantly, how to go about appraising and amending such formal definitions. This inability to produce and evaluate a consensus formalism may be caused by afflictions such as blinding by downward seepage or theoretical amnesia. Its consequence is that some of the fundamental constructs or propositions that drive scientific inquiries, the things that cause reality to be observed by the empirical scientist through a very specific set of tainted glasses, are never evaluated for their verisimilitude.

The Existential Question

One very important theoretical construct that manages to escape evaluation is probability, or more accurately: The probability theorems and derived statistical properties that are assumed to apply to the physical systems in which we may observe psychological phenomena. In psychological science it is generally assumed that the phenomena it studies obey the same rules as the phenomena that may be observed in a classical ergodic system (e.g., Molenaar, 2008; Molenaar & Campbell, 2009). Without going into too much detail, the classical ergodic condition assumes the space averaged behaviour of a system is equal (in the limit) to the time averaged behaviour of the system. If the ergodic condition applies, the outcome of the following 2 experiments is the same: Throwing 100 dice on one occasion (cf. data points are measurements obtained from a sample of 100 participants) and throwing 1 die 100 times in a row (cf. data points are 100 repeated measurements obtained within 1 participant). If you disagree with this assumption when applied to human behaviour (and you should) you can probably begin to see the contours of the real crisis that needs to be resolved in the empirical social sciences.

The history of science reveals the importance of questioning whether sets of axioms or formalism still suffice to describe the phenomena in the empirical record. The period of 1925-1935 was a transitional period in science for many other reasons than the “invention” of inferential statistics. For example, the realisation by Born, Heisenberg and Jordan (1926) that a new conception of probability and stochastic events was necessary to describe the statistical properties of quantum systems was one of the building blocks that led to a permanent schism between the ‘harder’ and ‘softer’ fields of science. The insight was that a non-commutative probability theory was needed in which the sequence of occurrence of stochastic events (i.e., the order of multiplication of probabilities) will yield different outcomes. Incidentally, the applicability of non-commutative probability theory to describe the statistics of various psychological phenomena has gained some interested recently (e.g., Busemeyer et al., 2011).

According to Max Born, their discoveries were driven by a crisis in particle physics. There are some similarities between the crisis in physics and the one in psychology of the same era, both concerned problems with epistemology and ontology of competing theoretical perspectives. Several simultaneously existing theories about different appearances of reality based on conflicting ontology (e.g., waves and particles), were eventually integrated into a single underlying formalism. The quantum formalism is based on about four foundational works that define the nature of quantum systems, relevant phenomena and levels of analysis, thus including rules for statistical inference (Born et al., 1926; Dirac, 1930; von Neumann, 1932; Schrödinger, 1935). The basic quantum formalism doesn’t make any initial ontological claims, it lays down the general principles from which scientific endeavours should depart. It provides an arena in which theories can compete for precision and accuracy of their predictions.

Among other things, establishing a consensus formalism allows for a program of theory evaluation by strong inference (e.g., Platt, 1964). There will be no confusion about what counts as a phenomenon that belongs to the domain of reality described by the formalism or not, something that is much needed in psychological science. For theories operating within the formalism, there is no need for preregistration of hypotheses or a discussion about how a failed replication should be interpreted. Data and theory truly speak more for themselves and any peer can in principle check the derivation of a prediction from the theory and assess which measurement outcomes can be expected and what the most likely interpretation of the possible outcomes should be in terms of verisimilitude of the theory. Interpretations in terms of explanation, or, description of why the universe at the quantum scale behaves as the theories predict it does are a matter of great debate (e.g., more than a dozen interpretations exist). This kind of interpretation has however been completely irrelevant for the advancement of quantum physics. No matter which interpretation you prefer, the level of precision and accuracy of quantum theories is unaltered: They are the most precise and accurate theories about reality ever produced by science.

A Game of Clones: Never Let a Good Crisis Go to Waste

An existential question for the empirical social sciences now emerges, one that requires examining how much blinding and amnesia have affected contemporary scientific practice and the loss of confidence in the empirical record: Is the goal of the scientific endeavour to produce theories that are precise and accurate in the empirical sense, or should theories be produced with high explanatory power, whose ontology its practitioners can be realist about, like good radical empiricists? Perhaps the best description of the mindset that will be needed to answer the realist question, one that will prevent us from cloning and implementing the failed solutions of the past to resolve the recurring crisis, is provided by Henry Poincaré:

> ``Whether the ether exists or not matters little - let us leave that to the metaphysicians; what is essential for us is, that everything happens as if it existed, and that this hypothesis is found to be suitable for the explanation of phenomena. After all, have we any other reason for believing in the existence of material objects? That, too, is only a convenient hypothesis; only, it will never cease to be so, while some day, no doubt, the ether will be thrown aside as useless.'' (Poincaré, 1889
> 1905, p. 211).

This quote contains a clue that should raise some sympathy from scientists in other disciplines for the dire position the social sciences are in today. About the existence of material objects Poincaré notes that this, like the Æther, is just a convenient hypothesis about the constituents of reality. However, it is one that "will never cease to be so". Why should material objects receive this special status? Material objects are sensory phenomena, manifolds of immediate sensory experiences, or direct experience. The luminiferous Æther on the other hand was one of the invisible constituents of reality physicists needed to imagine into existence, because everything appeared to happen as if it were really there. It was like the fifth element, a quintessence to complete our understanding of the world around us. In a way, it is 'easy' for physicists to identify something like the luminiferous Æther as a theoretical concept that may be discarded. Perhaps the greatest advances in physics have been made after the scales of measurement depleted or inflated beyond the capacity of our senses to register them as direct experience. For psychological science, the classical analogue to the Æther as a quintessence would be the Spirit, or Soul, but what about more contemporary ones such as the information processing Mind, the attentional spotlight, or the input-output model of perception and action?

The task at hand for a science that studies phenomena of the mind is to detach theoretical propositions about manifolds of sensory experiences from beliefs in their existence as a fundamental constituent of reality, let's indeed leave that to the metaphysicians. One should realise that when a theoretical proposition is discarded, thrown away as useless, this is the throwing away of a previously convenient hypothesis, which means that science advanced and more of the unobservable structure of reality has bee uncovered. Physicists know the theoretical constructs allowed for by the quantum formalism, such as quarks and bosons will eventually suffer the same fate as the luminiferous Æther, but for now, everything seems to happen as if it they exist and competing constructs such as cosmic strings, though intuitively appealing and formally plausible, cannot be empirically appraised for their verisimilitude. Realism in science should concern the structure of reality that highly corroborated seem theoretical constructs seem to capture (cf. Structural Realism, Worrall, 1989), not their literal interpretation or one-shot observation.

This (hi)story deserves a positive finale and it can be found in the realisation that the successful resolution of the crisis in physics that resulted in the quantum formalism was the result of intense debates by the leading scientists during the early Solvay conferences. These scholars recruited and created the mathematical tools and formal language that was necessary to describe the phenomena that were observed in labs around the world. The consensus was about the formal description of these phenomena, not about their interpretation (see Einstein's famous quote about some dice and a deity). Although not currently interpreted as such, I see there is a consensus about the nature of empirical psychological phenomena as well. What is observed in psychology labs around the world is that some phenomena are difficult to replicate, that there may be phenomena that are exquisitely context sensitive, whereas others are extremely resistant to perturbation. Even those robust phenomena are elusive to exact numerical replication based on current theories. I believe a consensus will emerge about the fact that many relevant phenomena currently in the empirical record, are insufficiently described as observations originating from an ergodic system.

These problems are solvable, but this will require recruiting and possibly creating mathematical tools and formal language. The task is to find a way to collectively agree on this solution, we need a 21st century analog to the Solvay conferences and I know just the right venue: The open science community. The movement harbours all the potential to finally make a difference and stop playing this tedious game of cloned 'solutions'. If there ever were a time in which the social sciences should attempt to join the tapestry weaving guild, it is right here, right now! We are witnessing the emergence of a science community that is ready to be open and reproducible. It has already initiated a close inspection of the patch- work quilt by questioning the veracity of theoretical claims in journals, on open fora and the scientific blogosphere. Increasing numbers of community members are engaging in post-publication peer-review, see Had I Been A Reviewer and the blogs linked there, or the PubMed Commons commentary system, the Winnower . The reason these developments are taking place right now is that the stamp of approval provided by a consensus among just a few expert peer reviewers is no longer accepted as sufficient for corroboration of a theory or hypothesis (see e.g., Van Noorden, 2013). What is needed is a consensus of a majority of the scientific community and online platforms are being developed to facilitate this process.

In 1935, Jastrow shared the wish to establish a sane science of psychology:

> ``There are consoling reflections. A science that can endure the ravages of two such distempers as behaviorism and psychoanalysis and recover without permanent disfigurement must have a lusty constitution. Still more, when I dwell upon the rich heritage of supremely significant knowledge which is all entitled to be called psychology, and the vitality of the tasks awaiting the psychologists of the future, the winter of my discontent becomes tinged with the promise of a glorious summer, when all psychologists shall practise the sanity they preach.''

I can feel it: **Summer is coming…** let's make it a glorious one, or at least a little more sane.

Fred Hasselman
Lost Archetype Labs, Lent
March 2014

ontology

noun on·tol·o·gy \än-ˈtä-lə-jē\

Definition of ONTOLOGY

1: a branch of metaphysics concerned with the nature and relations of being

2: a particular theory about the nature of being or the kinds of things that have existence

http://www.merriam-webster.com/dictionary/ontology

Das Wesen
das begriffen werden kann
Ist nicht das Wesen
des Unbegreiflichen.

- Lao Tzu (±500 BCE, Tao-Te-Ching)

In scientific discourse, there is often much confusion about ontology and epistemology, as well as sufficiency and necessity.

Apocrypha

LAO-TSE - TAO TE KING Das Buch vom Weltgesetz und seinem Wirken Verlag Otto Wilhelm Barth, Erstauflage 1928.

Notes

Chapter 1: Theoretical Diversity

1. This chapter may be cited as:

   Hasselman, F. (2014): Dealing with theoretical diversity: Ætiologies of developmental dyslexia as a case in point. In *Beyond the boundary. An analysis of verisimilitude and causal ontology of scientific claims: Ætiologies of developmental dyslexia as a case in point.* figs**hare**. http://dx.doi.org/10.6084/m9.figshare.1064427

2. The supplemental materials to this chapter may be cited as:
   Hasselman, F., (2014). Beyond The Boundary - 2nd Chapter: Supplemental Materials. Retrieved from Open Science Framework, https://osf.io/9mtpq

Chapter 1

Dealing with Theoretical Diversity

Ætiologies of Developmental Dyslexia as a Case in Point

1.1 The elusive ætiology of developmental dyslexia

When a child fails to acquire a proficient level of reading and spelling performance after about two years of regular literacy education in their native language and such possible causes of delay as below average intelligence, a specific sensory or motor deficit, or a general learning disability have been ruled out, the diagnosis is usually developmental dyslexia. Though many dyslexic readers will benefit from a timely intervention, the impairment will persist into adulthood with varying severity. Estimates of people afflicted with this specific learning disorder vary from country to country and ranges from 5-20% of the population (Blomert, 2005).

The definition of developmental dyslexia provided above is based on exclusion criteria. It describes the inability of otherwise typically developing children to acquire a proficient level of reading and spelling after regular literacy education (here typical just means average or normal with respect to development). Of course, there are more sophisticated diagnostic definitions that include a genetic component and mention a neurological basis for the impairment (cf. Fletcher & Lyon, 2008). There are also many more exclusion criteria based on results from elaborate test batteries and comorbidity with, for example, ADHD or Autism (e.g., Blomert, 2005). All these definitions remain however descriptive (or statistical) in nature and are essentially based on exclusion criteria. The practice of including genes or a brain region or function as an essential part of the definition appears to provide more precision or a deeper understanding of what causes developmental dyslexia, but that is just illusory precision. To date, there is no specific genetic profile (cf. Grigorenko, 2001), no specific brain structure or function (Caylak, 2009; Eckert, 2004, 2010) and no specific psychometric test profile (Blomert & Vaessen, 2009; Ramus & Szenkovits, 2008; Wimmer & Schurz, 2010) that can serve as the ultimate diagnostic tool. So what is it that is so specific about this specific learning disability, other than the observed problems with acquisition of proficient reading and spelling ability?

A quick survey of recent literature pertaining to the performance of dyslexic readers on a wide variety of tasks and experimental conditions seems to suggest that there is always some sample of dyslexic readers in some part of the world that can be found to deviate from average performance. There now exist a plethora of theoretical accounts of developmental dyslexia that explain such observed deviations from average performance as impairments in low-level sensorimotor processes or high-level cognitive processes, or both. The reported deficits span almost **all modalities of perceptio**n (Beaton, Edwards, & Peggie, 2006; Breznitz, 2003; Goswami, Fosker, Huss, Mead, & Szűcs, 2010; Huss, Verney, Fosker, Mead, & Goswami, 2010; Skoyles, 2004; Talcott et al., 2003; Tallal, 2004), include **deficits in motor contro**l (McPhillips & Jordan-Black, 2007; Nicolson & Fawcett, 2006; Ramus, Pidgeon, & Frith, 2003; Savage, 2004), **balance** (Rochelle & Talcott, 2006; Stoodley, Fawcett, Nicolson, & Stein, 2005), **attention deficits** (M. Reynolds & Besner, 2006; S. E. Shaywitz & B. A. Shaywitz, 2008; Valdois, Bosse, & Tainturier, 2004), **impaired cognitive abilities** (Aleci, Piana, Piccoli, & Bertolini, 2010; Heim et al., 2008; Helland, 2007), **fluency of naming** (Araujo, Pacheco, Faisca, Petersson, & Reis, 2010; Vaessen, Gerretsen, & Blomert, 2009), **learning** (Menghini, Vicari, Mandolesi, & Petrosini, 2011; Nicolson, Fawcett, Brookes, & Needle, 2010; Vicari et al., 2005) and **language** (Berninger, 2000; Joanisse, Manis, Keating, & Seidenberg, 2000; Koster et al., 2005). The studies reported here represent just a small anthology of the literature; the actual number of deficits proposed by scientists is much larger.

Areas of inquiry that have received much interest in the past decades due to their perceived potential to shed more light on the underlying causes of the observed impairments are the genes and brains of dyslexic readers. However, as mentioned above, the current empirical record of neurobiological facts and neural and genetic correlates of behaviour related to developmental dyslexia has not been decisive in the resolution of any theoretical dispute that existed before in vivo brain-imaging and gene-sequencing became available as tools for scientists. It has not provided a consensus on the aetiology of the reading impairment (Dowker, 2006; Marinelli, Angelelli, Di Filippo, & Zoccolotti, 2011; Pugh et al., 2001; Ramus, White, & Frith, 2006; Ramus, 2003a, 2004; Stanovich, 1985, 1988), neurobiological evidence has not given more scientific credibility to one theory of (impaired) reading and spelling over others (Heath, Bishop, Hogben, & Roach, 2006; Howes, Bigler, Burlingame, & Lawson, 2003; Mechelli, Gorno-Tempini, & Price, 2003; Ramus & Szenkovits, 2008; Ramus et al., 2003) and it has certainly not helped with achieving a clearer definition of developmental dyslexia as a specific learning disability or developmental psychopathology (Bishop & Snowling, 2004; Fletcher & Lyon, 2008; Frith, 1999; Landerl & Wimmer, 2000; Lyon, 1995; Lyon, S. E. Shaywitz, & B. A. Shaywitz, 2003; Wolf, 1999). In fact, new disputes about the'real' neural correlates of impaired reading have erupted, about the myth of the existence of a visual word-form area (Price & Devlin, 2003), the significance of the observed cerebellar dysfunction (Nicolson & Fawcett, 2006) and how letters and speech sounds are integrated in the brain (Blau et al., 2010; Blomert, 2010). The main difference with the times of Galaburda and Kemper, who in 1979 presented some of the first evidence for anomalous neural organisation in a post-mortem study of the brain of a dyslexic reader, appears to be that now each hypothesised deficit comes with its own neural correlates. Thus providing an apparent existence-proof for the deficit in question. Not surprisingly, most suggested deficits have their own treatment program (see section 1.2.3) and new intervention studies emerge based on studies of brain activity as well (Breteler, Arns, Peters, Giepmans, & Verhoeven, 2010). It is even the case for some hypothesized deficits to have their own associated genetic correlates (Grigorenko, 2001; Ramus, 2004).

How can the difficulties with the acquisition of proficient reading and spelling ability give rise to such diversity in apparently veridical ætiologies proposed by scientists? That is, explanations of causes of the impairment all seem to be evidenced by a considerable empirical record spanning every level of analysis from genes to overt behaviour. Fletcher (2009) interprets the current status quo as the result of the evolution of developmental dyslexia as a scientific concept. After decades of scientific studies some authors are indeed questioning whether the proposed deficits are'real' (Moores, 2004; Ramus & Szenkovits, 2008) or whether dyslexia actually exists as a collection of deficits (Elliott & Gibbs, 2008). In any case, history may be repeating itself, since summarizing papers with a title along the lines of'What have we learned so far?' can be found in almost every decade (Hudson, High, & Al Otaiba, 2007; Snowling, 1996; Stanovich, 1985; Vellutino & Scanlon, 1998; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Most scholars and research programs are well aware of all the other theories out there, but they still embark on a quest for the Holy Grail: To identify a single, one-way, causal pathway from genes to impaired reading. The type of causality the scientists pursue in their quest is a chain of efficient causes hypothesised to be attributable to components or component processes at various levels of analysis as depicted in Figure 1.1. An example of such a quest is a series of papers by Ramus (2003b, 2003c, 2004), who concluded after an

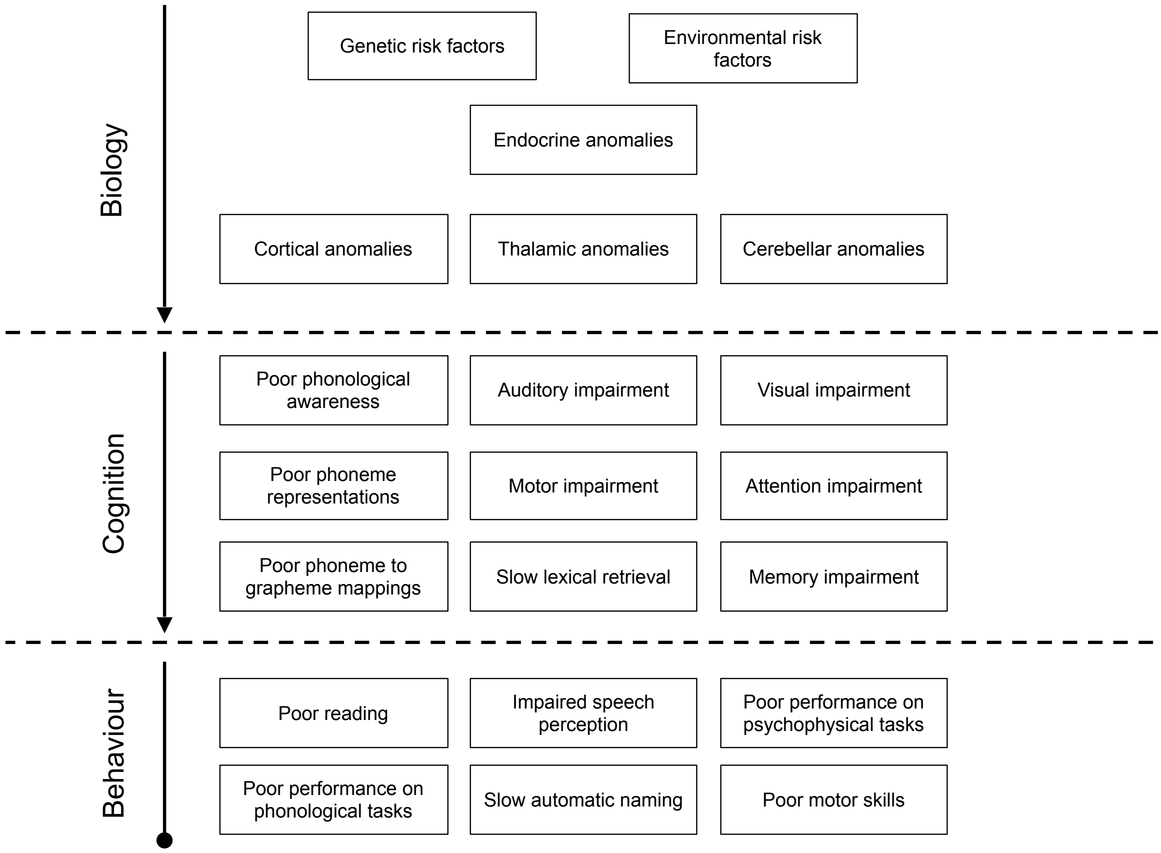


Figure 1.1 – The quest for the Holy Grail in the scientific study of developmental dyslexia: Connect the dots and find the proper arrangement of components from biology to behaviour that cause developmental dyslexia.

extensive review of the literature that a whole range of proposed deficits may be associated with developmental dyslexia, but play no causal role in its ætiology. In other words, those deficits should be considered epiphenomena of the'true' cause.

Ramus' quest in that series of articles was to draw causal pathways through the phenomena (the boxes in Figure 1.1) that were substantially backed by empirical results. It is exactly this type of scientific inference, in combination with the inconclusive results of such inferences that –from a meta-theoretical[1] point of view– raises the question whether the proposed causal mechanisms should be considered to be similar (high *similitude* of theories). Or, perhaps they have a low'truth-likeness' (low *verisimilitude*), or at least are not appraised in a rigorous fashion to assess their verisimilitude. I take that it is uncontroversial to suggest that the purpose of a science is to evaluate whether the claims it produces about the way the world works have some truth to them and eventually select the most truthful of all such claims[2] In the remainder of this chapter I will attempt to analyse the origins of the theoretical diversity and the apparent incapacity of the scientific method to resolve it by means of appraising verisimilitude of individual claims. The conjecture I have made my goal to elucidate in this chapter is: *A theoretical account about the ætiology of developmental dyslexia of higher verisimilitude than current accounts, should be able to explain how it is possible that the current plenitude of theories can be hypothetically true at the same point in time* (or space for that

[1] Meta-theory is the empirical study of scientific theorising (Meehl, 1992; 2002; 2004a).

[2] I will use the term 'truth' here in the most practical sense possible: To distinguish between scientific claims, that is, the precision and accuracy with which they describe observable phenomena in some domain of reality. If 'truth' is an uncomfortable word in this context, I see no objection to exchange it for 'scientific credibility'. There is however a formal distinction between a hypothetico-deductive test and a crediballistically-deductive test of scientific claims (see Rozeboom, 1982).

matter, see the recent special issue of Dyslexia 2013, volume 16, issues 3 and 4: Investigating the links between Neurocognitive Functions and Dyslexia).

It is thus explicitly not the purpose of this chapter, or any other chapter in this book, to describe a successful scientific quest that ends in the discovery of the Holy Grail: A unique causal pathway. An empirical quest will be reported in the chapters that follow, but its Holy Grail takes on a very different shape and form. The results of the (meta-)theoretical analysis and historical and philosophical perspectives on the conjecture-made-goal described above, have motivated virtually all of the theoretical and empirical decisions made in studies described in chapters 2-5.[3] The present chapter is an essential general introduction to the more specific empirical inquiries into the role of speech perception in developmental dyslexia.

1.2 Corroborative evidence: A growing body, or morbid obesitas?

The introduction to the scientific study of developmental dyslexia so far, requires from me to evidence at least two claims and offer some explanations if those claims turn out to be true:

Claim 1 The number of substantive[4] theoretical accounts for the ætiology of developmental dyslexia has been growing rather than shrinking.

Claim 2 High quality data measured to corroborate one theoretical account and its causal pathway is of no consequence for the perceived verisimilitude of a competing theory.[5]

To do so I present a historical analysis of the terminology used in the scientific literature on the ætiology of developmental dyslexia.

What exactly causes developmental dyslexia has been debated for quite some time, at least since competing theories appeared in the literature in the 1970s. These theories disputed whether the reading deficit was caused by impaired auditory processing or impaired phonological processing (Bradley & Bryant, 1978; Tallal & Piercy, 1973) and were further developed based on evidence from post-mortem studies of the brains of dyslexic readers (Galaburda & Kemper, 1979; Galaburda, LoTurco, Ramus, Fitch, & Rosen, 2006). Both assumed an impaired theoretical entity, the phoneme representation, to be causally entailed in causing the difficulties with acquiring a proficient level of reading and spelling. What may be reasonably expected of a discipline of science is that it works towards a more unified formal description of reality instead of generating more contradictory ones. Figure **1.2** reveals, among other things, that the latter situation seems likely when scientific descriptions of the ætiology of developmental dyslexia are concerned.

The graphs in the figure are representations of the connectivity between terms (categories of words) found in 1407 abstracts of scientific papers on the ætiology of developmental dyslexia on PubMed (see Appendix A for details). The abstracts were analysed as a corpus based on the decade of their publication date. As an example, take the 1970-1980 era. The boundaries of the graph are inhabited by a few terms that are closely associated. The upper side contains auditory perception (per~aud), pathology in brain structure (cns~str~, cns~pat), treatment (~trt~), motor learning and speech production (lrn~mot, ~spr~). These terms and associations seem to evidence studies about

[3] Pardon my hindsight bias. Chronologically, this chapter was completed last, but note that five years ago, I could not have written it.

[4] I will use 'substantive theory' almost exclusively to refer to a theory backed by a substantial amount or quantity of empirical evidence. This is somewhat different from what Paul Meehl described as'The theory has money in the bank'. The latter refers to the track record of corroboration events that constitute risky predictions by a theory and it is questionable whether the predictions by the theories I will discuss can be categorised as risky. Or corroborated for that matter.

[5] Stated less formally, there appear to be no theories with silverback-alpha-male-pack-leader-results, to which theories competing for resources and proliferation succumb (empirical grid-lock). Or, perhaps there are only alpha theories; all Indians, no chiefs, and this is a stand-off (ontological indifference).

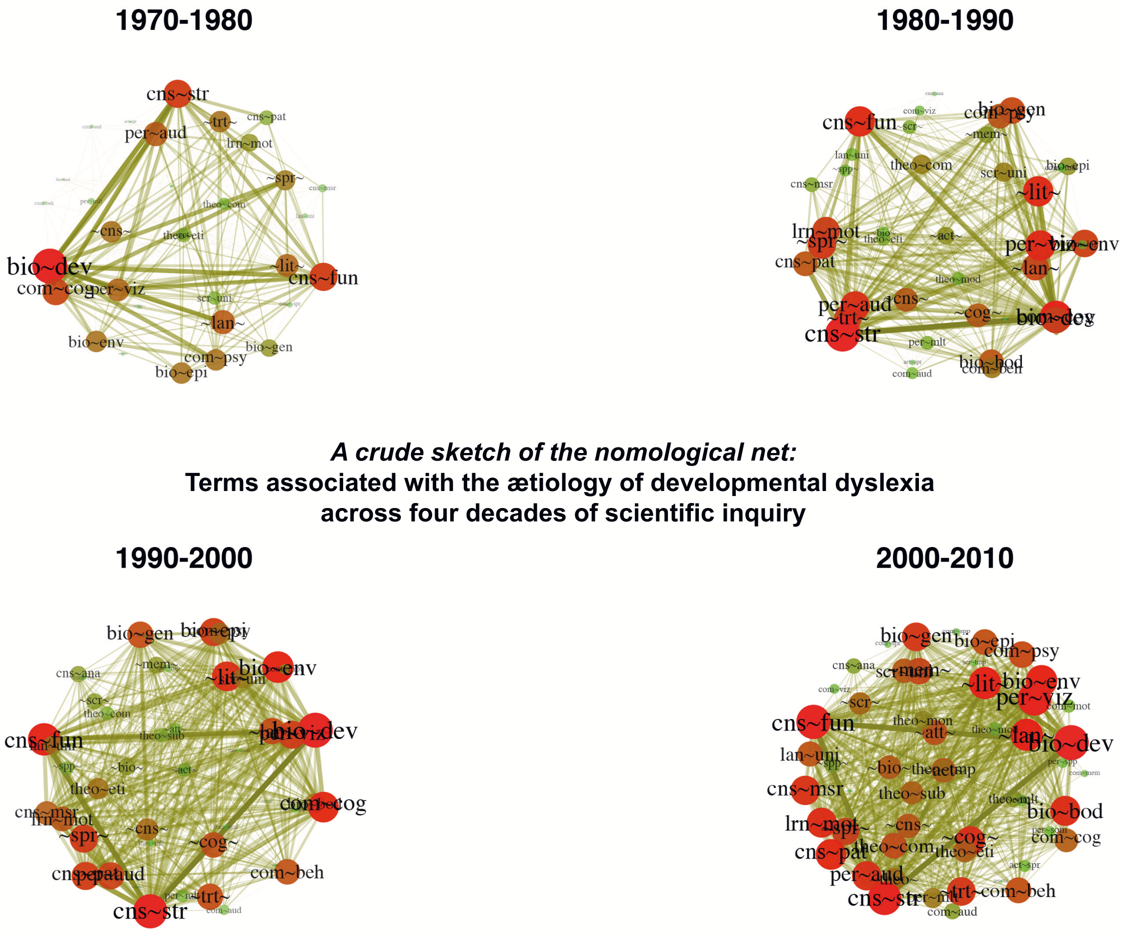


Figure 1.2 – Weighted graphs of term associations used in four decades of the scientific study of the causes of developmental dyslexia. The associations represent a cumulative sum of terms occurring in the same abstract of a scientific paper (*N* = 1407). The abstracts were obtained via a PubMed search query on the ætiology of developmental dyslexia. The large, red nodes have a high (standardised) degree, that is, more connections to other nodes in the network (see text and Appendix A.1 for details).

the impaired auditory processing hypothesis mentioned earlier (Tallal & Piercy, 1973). The other two clusters contain terms like language and literacy (~lan~, ~lit~) associated with brain function (cns ~fun) and this seems in accordance with the impaired language processing account mentioned earlier (Bradley & Bryant, 1978). There are also many connections to comorbid diagnoses, cognitive or psychopathological, associated with biological terms including development (com~cog, com~psy, bio ~dev, bio~env) and visual perception (per~viz). No doubt these terms appear due to attempts to define and diagnose dyslexia, but they also herald the arrival of the (visual) magnocellular deficit hypothesis (cf. Stein, 2001).

The graphs reveal that the structure of the graphs does not change just due to an increasing number of terms (the nodes or vertices), but it is clearly the case that the terms become less uniquely specified. That is, the number of connections (edges) a node has to other nodes in the network increases and nodes representing new terms do not lag behind'old' ones in this respect. It really seems to be the case that what started with roughly two to three substantive theoretical accounts has grown into a vast collection of different theories, deficits and hypotheses, all backed by a considerable empirical record that contains every type of data from behavioural, neurophysiological, to genetic. Also, every design seems to be represented, there are data sets that were acquired experimentally, by means of decade long prospective studies and/or by comparing subpopulations, as well as computer simulations. It is truly an impressive body of evidence, but what does it evidence?

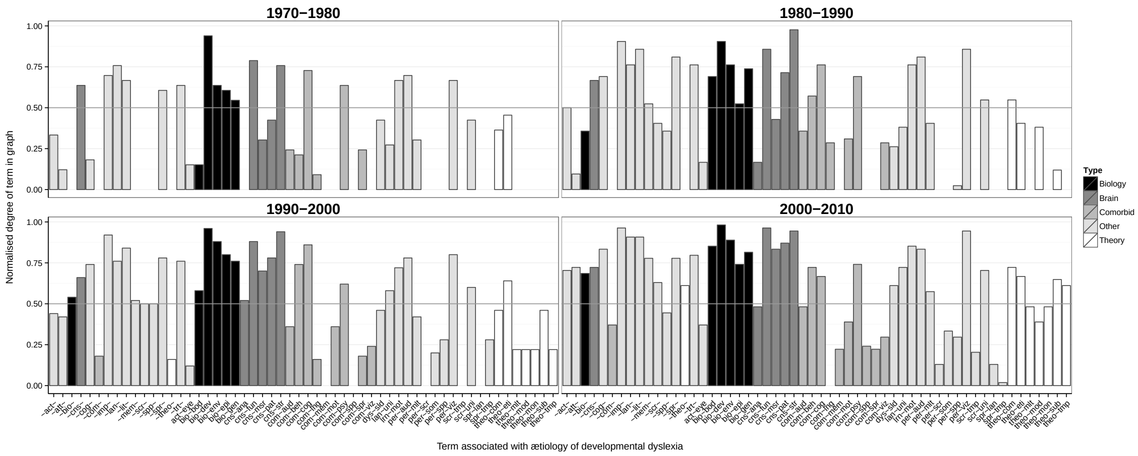


Figure 1.3 – Bars represent the normalised degree distribution for each node in the graphs of Figure **1.2**. This represents the number of connections each term has to other terms divided by the total number of connections in the graph. The bars are colour coded to indicate terms that can grossly be categorised as Biology, Brain, Comorbidity and Theory (compare to levels of Figure **1.2**). See text and Appendix A for details, such as, a list of the terms and their meaning in Table A.1.

The pattern inferred from eyeballing the term networks can be quantified by calculating the degree distribution in the network. Figure **1.3** represents for each term the normalised degree of the node. A normalised degree of 0.5 (the horizontal line) means that the connections of a node to other nodes consist of 50% of the unique connections that are available in the network. This does not imply that a node is particularly important; in a fully connected network all the nodes host 100% of the connections. Interestingly, more varieties of comorbid diagnoses are mentioned in association with other terms as time goes by. This could be related to the observation in the first paragraph that a definition of developmental dyslexia is still mostly a definition of exclusion criteria, like excluding comorbid diagnoses. Another pattern that emerges is that as more terms appear, they are often connected to more than 50% of the other terms in the network. Terms categorised as Biology (e.g., genes, epigenetics) and Brain seem to‘lead the way’ in reaching the 50% mark, they are almost always mentioned in association with any of the other terms in the network.

1.2.1 The nomological network: A rough sketch

To interpret what the degree signifies, it is important to consider what kind of network this is. What do its nodes and edges represent? I believe it is sensible to argue that the networks represent a very crude sketch of a nomological net as suggested by Crohnbach and Meehl (1955). They described nodes of a nomological net representing theoretical (or ontological) entities and their connections to other entities as lawful relations (functional or compositional). Using the terminology of graph theory and complex networks in combination with basic meta-theoretical concepts, I suggest a more detailed specification of the nomological net is possible. The formalized theories of physics will often consist of deduced entities and laws, whereas theoretical accounts in social science are predominantly based on induction by statistical regularities (see Table **1.1** for a reminder of the differences between deduction and induction, cf. Salmon, 1999). If laws were allowed to be either universal or statistical and the entities can pertain to particular facts or general regularities, the four types of scientific explanation in the received view of scientific explanation (Hempel & Oppenheim, 1948; Hempel, 1968) could be represented by the net (see Box **1.2**). By using a weighted and directed graph, one could indicate Deductive-Nomological / Deductive-Statistical explanations as directed edges (one way connections) and Inductive-Statistical explanations by equating the degree of inductive strength to the edge weight.

Box 1.1: The Differences Between Deduction and Induction. Adapted from Salmon (1999, p.11)

DEDUCTION	**INDUCTION**
All humans are mortal	
Socrates is human	All observed ravens have been black
Socrates is mortal	*All ravens are black*
1. In a valid deductive argument, all of the content of the conclusion is present, at least implicitly, in the premises. Deduction is *non-ampliative*.	1. Induction is *ampliative*. The conclusion of an inductive argument has content that goes beyond the content of its premises.
2. If the premises are true, the conclusion must be true. Valid deduction is necessarily *truth preserving*.	2. A correct inductive argument may have true premises and a false conclusion. Induction is not necessarily truth preserving.
3. If new premises are added to a valid deductive argument (and none of the original premises is changed or deleted) the argument remains valid. Deduction is*erosion-proof*	3. New premises may completely undermine a strong inductive argument. Induction is *not erosion-proof*
4. Deductive validity is an *all-or-nothing* matter; validity does not come in degrees. An argument is totally valid or it is invalid.	4. Inductive arguments come in different *degrees of strength*. In some inductions the premises support the conclusions more strongly than in others.

An ideal rendering of a nomological net should thus be able to represent a topology of scientific explanation: A single theory, a larger theoretical framework, or all the different theories that can be connected in some way. Only a subset of the net will make contact with the empirical record, that is. The verisimilitude[6] of purely theoretical entities could be the node degree or if observable entities are associated to it statistically, the sum of the edge weights could be used. The structure of the topology can be thought of to represent the'success' of scientific theorising, or its global verisimilitude. The goal is to strengthen the logical structure of the net by strengthening connections in a derivation chain upon corroboration or deduction, but pruning derivation chains that fail severe testing are logically inconsistent. Graph theory and network analysis also allow detecting similitude of graphs (isomorphism) this offers a potential tool to examine whether unification is possible.

The networks under consideration here do not represent this ideal rendering of the net. The nodes could represent theoretical constructs and perhaps the edges even laws, but not as just described. First, it is important to remember that the terms in the networks are conditional on the result of a search query that contained terms used to refer to developmental dyslexia and its ætiology (otherwise a much larger set of abstracts would have been found). The nodes representing the search terms (dyslexia, etc.) have been deleted, as we already know they appear in each abstract and are therefore connected to all the other terms. Otherwise each hub in the network would be connected to nearly all other hubs due to these terms. Second, the terms that make up the nodes are general categories, not strictly representing theoretical entities as some words just mean different things in different contexts. Mentioning a brain structure in an abstract could have theoretical

[6]Yes, objections were raised against the received view, most importantly that the goal of scientific explanation is unification, not causality (Kitcher, 1989). In the everyday practice of empirical science, causality will be on the mind of the researcher, not unification. The received view seems to me a most complete account of scientific explanation that strikes a balance between idealised science and a taxonomy of theorising in the wild. Moreover, unification is explicitly defined as a goal of science in the notion of strengthening the logical structure of the net. Also, the received view has been declared a straw man of philosophy of science and has been defended quite successfully recently (see Lutz, 2012).

Box 1.2: The Four Forms of Scientific Explanation According to the Received View. Adapted from Salmon (1999, p.16). Note that the Nomological Net Metaphor implies the Goal of Unification.

	Explananda	
Laws	*Particular Facts (Ideographic)*	*General Regularities (Nomothetic)*
Universal laws	D-N Deductive-Nomological	D-N Deductive-Nomological
Statistical laws	I-S Inductive-Statistical	D-S Deductive-Statistical

reasons as well as be the result of an empirical study of that structure.

Finally, the weight of the edges connecting the nodes in the graphs is calculated as a sum over the abstracts, indicating how often the term pairs were mentioned in concert in the corpus. These weights could very well evidence an underlying empirical law, as the terms must be correlated in some way, but this cannot be interpreted as a correlational measure. As for the structure of the network topology, I see no reason why the general idea of the structure of scientific explanation represented as entities that can be related either by frequency of co-occurrence or by the number of connections they make to other terms does not apply here. This is in fact how graph theory was used recently to analyse the structure of symptoms of psychopathology listed in the DSM-IV (Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011).

Even when these restrictions on the scope of the results are taken into account, from the perspective of theory evaluation as an attempt to strengthen those structures in the net that are scientifically credible, and at the same time prune the less credible laws and their associated nodes, the graphs do not show a tendency towards less connected nodes, the opposite is the case. When each term ends up associated to every other term, the network cannot be very informative as a knowledge base. Lebel and Peters (2011), discussing claims of evidence for paranormal phenomena (Bem, 2011), suggested that a weak knowledge system contributes to the publication of such incredible claims in psychological science. The verisimilitude of the entities and laws involved in a theoretical claim should be severely tested and ruthlessly discarded if they cannot pass the tests. That is how the logical structure of knowledge represented in the nomological net increases. In addition to (or instead of) a possible low logical structure of the theories represented by the networks in Figure **1.2**, one could interpret the observed non-specificity as the sign of a high similitude of the theories that spawned the studies whose abstracts were analysed.

This could be an explanation for *Claim 2*: Empirical evidence cannot be decisive in assessing the verisimilitude of competing theoretical claims, not because they are not severely tested, but because the entities and laws from which observations are predicted are essentially two sides of the same structural coin. The question of similitude will be discussed in the last paragraph of this chapter.

1.2.2 Meet ~theo~ … there goes the neighbourhood!

The final support for*Claim 1* (the number of substantive theoretical accounts has been growing rather than shrinking) and its logical consequence Claim 2 (the credibility of substantive theoretical accounts is invulnerable to apparent corroborations of competing substantive theoretical accounts) will be provided by examining the category Theory. The bars representing different theories are grouped at the end of the x-axes in Figure **1.3**. These terms exclusively represent word combinations

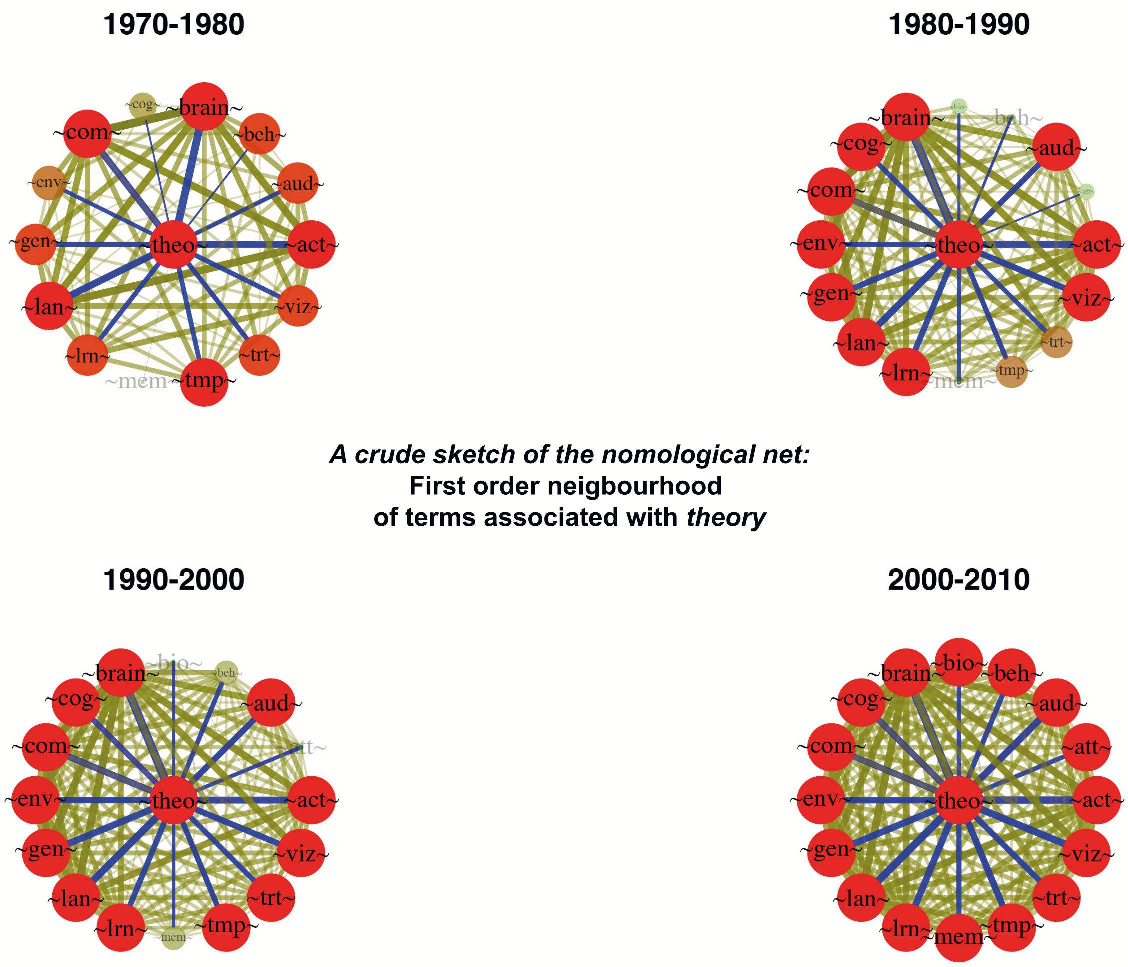


Figure 1.4 – The graphs represent the first order neighbourhood of the ~theo~node based on the networks displayed in Figure **1.3**. Here, specific classifications of terms were collapsed into broader categories. The nodes in the graphs now have an equal number of categories for each decade. The blue connections represent direct associations of theory terms to other categories. The other connections represent associations between the terms in the first order neighbourhood of node ~theo~. See text and Appendix A.1 for details.

associated with a particular theoretical account, from multi-causal theories to those that focus on deficits in temporal processing (see Table A.1). One could object that their appearance is normal due to the progress of a scientific field of study and advancement of technology in general. The number of studies with downloadable abstracts published on the subject seems to grow as a power law of 2^n (n = 36, 123, 264, 686 respectively). The networks do reveal that more detailed, specific terms appear in the scientific record as time goes by. One could argue that as science advances and advanced technologies like in vivo brain imaging and behavioural-genetic analysis are more commonly available, this will boost connections between terms that were previously unconnected; a logical consequence of scientific progress. To gain more insight in the relevance of the evolution of this nomological net for appraising and amending theories about the ætiology of developmental dyslexia, the corpus of abstracts was re-analysed as shown in Figure **1.4**. The focus of the analysis is to assess the impact of the explosive growth of the empirical record on theory evaluation. To do so, all the specific terms were collapsed into 16 broader categories that occurred in all four decades and every word indicating a theoretical statement was collapsed to the term ~theo~(17 nodes, see Appendix A.2, Table A.2). As a next step the node representing theory terms and its direct neighbourhood was isolated from the total network. This is a neighbourhood of order 1, meaning it contains only nodes that have one direct connection to the theory node. Those connections are displayed in blue in Figure **1.4**, the other connections are those that exist between the other nodes in the neighbourhood.

Using the broader categories, the networks now appear to present a more balanced picture. In 1970-1980 on this much coarsely grained scale, terms directly associated to theory indicator words were also connected to each other. Important is to note the difference in the weights of the connections (how often did the terms appear together in an abstract in that decade?) and the degree of the nodes (how many connections does the node make to other nodes?). These values (represented by size and colour of nodes and edges) are rescaled to the minimum and maximum values observed in each decade and therefore comparable across decades. The conclusion must therefore be that the pattern persists: The degree distribution and weights of the connections of the nodes in these neighbourhood networks increase over time. One way to express this for an individual node is to look at its co-citation coupling (Small, 1973). Two connected nodes are co-cited if another node'cites' (connects to) both of them. For each of the four decades the average co-citation coupling of ~theo~to the other nodes is: 11.7 (*SD* = 1.9), 14.3 (*SD* = 1.1), 14.6 (*SD* = 0.7) and 15 (*SD* = 0), respectively. This means that in the last network where ~theo~is connected to ~gen~all the other nodes connect to both ~theo~and ~gen~. Such is the case for all possible node combinations with ~theo~as there is no dispersion in co-citations. A summary metric of the weights of the connections is usually expressed as the sum of all edge weights connecting to a node, or the graph strength (Barrat, Barthélemy, Pastor-Satorras, & Vespignani, 2004). For the ~theo~node the graph strength is: 24.4, 41.2, 52.7 and 69.0 for each consecutive decade.

To summarise: More connections emerge between theory words and other terms as time goes by and the strength of those connections (how often they co-occur in abstracts of scientific articles) also increases. In the most recent decade (2000-2010) the terms found in 686 abstracts that are most directly associated with theoretical claims about the ætiology of developmental dyslexia form a network in which:

1. Indicators of theoretical claims are associated to every other term in the network.
2. The strength of these associations is equal and maximal for all connections.
3. Each node that is associated to a theory word is also connected to every other node in the network.

1.2.3 The'inventors' in social science: Interventions as technology.

What are the causes and consequences of this weak knowledge system? One consequence that should raise some concern is the development of'technology' based on these ætiologies of developmental dyslexia. In the'applied' social sciences the scientific knowledge produced by the'pure' fields is turned either into diagnostic or performance measuring instruments or into intervention programs whose goal is behavioural change, which can be anything from optimal design of advertising campaigns to a treatment program for a psychopathological disorder.

In the case of developmental dyslexia, the technology produced is intervention programs that ultimately should ameliorate reading and spelling performance. The fact that just about any theoretical claim seems to enjoy empirical success has resulted in an equal plethora of intervention studies, some of which are so esoteric, or'armchair unlikely', that I consider it a real possibility that as a consequence of the inability of scientists in these fields to properly evaluate their theories for verisimilitude, the lives of children were adversely affected. At the very least it should raise questions about the ethics of intervention research. A whole range of interventions, some without any apparent relation to reading and spelling, continue to be proposed by serious scientists in serious scientific journals. The interventions reveal we are dealing with a science that is realist (in the most direct sense) about the phenomena and entities its theories posits to exist in reality. In principle, this is not at all a bad trait for an individual scientist, as long as there is the realisation that the ontology is a temporary one. In an arena with so many different competing theories claiming to'get it right', one would expect some more pervasive scepticism from a genuine scientific endeavour.

The proposed treatments are clearly inspired by the (causal) ontology associated with a specific deficit (see Figure **1.1**) and the rationale behind most of these interventions is not very different from attempting to repair a faulty component in a machine. Recent examples of studies reporting treatment effects are: Adding Fish Oil (Omega-3 fatty acids) to the diet to improve the diminished myelination of the *magnocellular part of the central nervous system* (e.g. Cyhlarova et al., 2007 » magnocellular deficit hypothesis); Using coloured lenses and coloured overlays to improve reading fluency (e.g. Lightstone, Lightstone, & Wilkins, 1999; Whiteley & Smith, 2001) » *visual disturbance / visual stress hypothesis*); Music therapy (Cogo-Moreira et al., 2012; e.g. Overy, 2003) » *rhythm / coordination imbalance hypothesis*); Intensive training with tonal sweeps and acoustically modified speech (e.g. Tallal, 2004 » *auditory temporal processing hypothesis*); Neuro-Feedback training (e.g. Breteler, Arns, Peters, Giepmans, & Verhoeven, 2010 » *brain dynamics deficit hypothesis*); Training motor skills (e.g. D. Reynolds, Nicolson, & Hambly, 2003; D. Reynolds & Nicolson, 2007 – *cerebellar deficit hypothesis*); Presenting stimuli to visual, auditory and touch modalities opposite a dysfunctional hemisphere (e.g. Smit-Glaudé, van Strien, Licht, & Bakker, 2005) » *hemispheric balance model of reading and dyslexia*); Training Rapid Serial Naming of pictures, colours, numbers and letters (e.g. Eleveld, 2005 – *rapid naming / fluency deficit hypothesis*); Training children by letting them play action video games (Franceschini et al., 2013) » *visuo-spatial crossmodal temporal attentional deficit / (dorsal) magnocellular deficit hypothesis*. These interventions were chosen because they appear rather unorthodox and they all but one report positive effects some of which are close to miraculous. Twelve hours of'training' with action video games was reported to cause gains in reading, spelling and phonological skills equivalent to 1 year of 'regular' remediation (Franceschini et al., 2013). There are many more reports of effective interventions in the literature, most of which are based on these 'regular', but in my opinion at least less harmful remediation programs: Repetition learning of reading and spelling performance.

As Novalis suggested in the epigraph, in social science, the inventors are also the true hypothesisers, they cast their nets and they always catch. This problem of'easy' evidencing of effects of intervention was recently discussed as one of the pervasive problems in psychological science (Boot, Simons, Stothart, & Stutts, 2013). The solutions offered mostly concern the design of experiments (e.g., active control groups as placebo stand-ins) and fail to recognise that interventions are in fact a technology produced by scientific theories and their tests should perhaps be treated differently from tests of theories (see Meehl, 1997). As has been known for a long time, effective technologies of the social and life sciences, from psychotherapeutic or surgical intervention to pharmacological treatments have a high likelihood of'breaking down' when they are brought into the real world where they have to be efficient as well as mere effective under ideal conditions of a clinical trial (cf. Cochrane, 1972).

My goal, as stated earlier, is to seek an explanation for the persistence of the weak knowledge structure in the empirical record, that is, focus on substantive theories and examine in what respect they are similar and if they are not, to what extent we can understand theoretical diversity arises due to weakness of applied methods and philosophy of science. First, I will sketch a metatheoretical account of the mechanism behind the erosion of authority of theory corroboration by empirical evidence (section 1.3). Second, I will examine the differences between theorising in the empirical social sciences and the natural sciences (section 1.4). Before solutions can be presented, I provide an overview of the problem of theoretical diversity and discuss some causes for its emergence specific for the object of study in psychological science (section 1.5). Section 1.6 is a proposal to define developmental dyslexia as a state-entity and study claims about its ætiology guided by degrees of theory specification. The remaining chapters in this book will be characterised according to the level of theory specification they set out to test.

1.3 Epistemic Sloughing: How to Erode the Authority of Corroborative Events

> ``If Psychology is ever to become anything more than a mere aggregation of opinions, it can only be by the establishment of some datum universally agreed to."
>
> -Herbert Spencer (1855, p. 8)

What appears to have occurred in the study of developmental dyslexia is that any significant pattern in the data was interpreted as a phenomenon of significance for the ætiology. Although direct tests of predictions by different theories do exist, the empirical results appear to have no authority on the level of perceived verisimilitude of the competing theories, even if those results pertain to empirical studies of the brain or genes and are therefore indirect discorroborations with respect to competing theories (Ramus, 2003a, 2003b, 2004; Ramus & Szenkovits, 2008; Ramus et al., 2006;). These theories are in *empirical gridlock*, or perhaps due to the shear volume of empirical phenomena that are perceived as being relevant, an *ontological indifference* has evolved among the researchers in the community. The energy spent on attempts to dispose of competing theoretical accounts is apparently perceived as a waste of time and resources by the field; in any case, such attempts do not seem to occur often. The result is a flooding of the empirical record with empirical facts representing unrefuted hypotheses, a condition I call Empiarrhhea.[7] Each phenomenon is'saved' to support a specific theoretical deficit and each deficit ends up with its own anatomical and functional brain anomaly and genetic anomaly. If there is anything evidence from neuroscience and genetics has done, it is an ontological truth status for different deficits that resonates more with a generally accepted world-view, than verisimilitude of a theoretical whole: Such deficits must exist as constituents of reality, after all, aren't we determined by our brains and genes according to science? This is why in Figure **1.1** the CNS and genes are at the top of the causal pathways, the origin of the derivative chain, with most (apparent) epistemic weight.

It seems as if the scientific method, the repeated application of the empirical cycle in which theories are tested based on their predictions of observable phenomena has turned into a vicious cycle of consisting only of corroboration events. I use the term corroboration event loosely to indicate that a theoretical prediction (and as such the whole of theoretical claims in the derivational chain) was 'evidenced' or 'confirmed' or 'supported' or 'verified' or 'not falsified' or 'considered plausible'[8] by an empirical observation according to a cultural convention specific to a community of scientists. Often this convention concerns turning the prediction into a hypothesis that is tested using an observational threshold for evidencing phenomena. The threshold represents a certain degree of confidence about the truth-status of the hypothesis, for example, whether the probability of an observation falls below the α-level (false positive rate) in null hypothesis significance testing (NHST). The 3σ evidence level and 5σ discovery level for elementary particles is used to decide how remarkable an observation is, compared with observations that may be expected if the predicted particle was not part of the description of reality (i.e., the standard model).

The corroboration strength, or weight of the event granting truth-likeness to a theory in the perception of scientists, should however not be confused with the magnitude of a probability or an effect size associated with the event. Together with a complex interaction of less objective judgements about the novelty, aesthetics, or risky-ness of the prediction and the severity of the test, it is the objective precision and accuracy of a theoretical prediction that decide the corroborative

[7] I introduced the term Empiarrhea to indicate the zealous empiricists practice of flooding the empirical record with cute facts that have no impact on the veracity of theoretical claims. Rick Dale (personal communication, 18-03-2011) pointed out one could also call this phenomenon Theorrhea indicating the influx of new theoretical accounts into the scientific record by theorists in search of a monism.

[8] These terms mean different things to philosophers of science and logicians. I hope my generalisation for the sake of the argument will be forgiven.

strength of the event and its effect on perceived verisimilitude of the theories involved (Mayo & Spanos, 2006; Meehl, 1990b, 2002). For example, in general a lesser corroboration strength is assigned to a convergence of evidence compared to the prediction of novel facts. General Relativity Theory (GRT) could account for Mercury's perihelion advance, a known anomaly to Newton's theory of celestial mechanics, but it was the corroboration event of a predicted observation (bending of star light during an eclipse in 1918) that convinced most scientists, and the general public (compare Hasselman, 2013 on using model fit indices as corroboration strength). One group of scientists formed an exception, most astrophysicists were more impressed by the after-the-fact prediction of Mercury's perihelion advance (Brush, 1989). Meehl (2002) offers an interpretation that I believe is crucial for understanding the weak structure of the nomological net described in the previous section. I will refer to it as the *epistemic sloughing*[9] of theoretical predictions by means of hypothesis testing. In order to understand the mechanisms of this process consider the different types of research that are commonly used in the empirical fields of the social and life sciences.

The kind of research conducted in the social and life sciences can be categorised (based on Lykken, 1968) as follows:

3. The effect of some treatment on some output variables in applied research (e.g., the testing of technology, (cf. Meehl, 1997). These studies are a special case of studies that:

   2. Examine the difference between two or more groups of individuals with respect to some variable. These studies are again a special case of studies that:

      1. Examine the relationship or correlation between two or more variables within some specified population.

In general, it can be said that data patterns that might evidence or refute hypotheses, irrespective of the kind of study that is conducted, will be patterns of association (e.g., variables share a degree of common variance, value sequences display a degree of recurrence or symmetry (breaking); Neyman, 1969). Whether or not an observed association between variables represents a causal entailment is a matter of interpretation of the results in the context of the study as a whole. At the most basic level, all studies are tests of sign predictions.

To predict the sign of a correlation is the least risky of predictions[10] a theory could produce about future empirical observations. The values can take on: Positive, negative, or no correlation, or in terms of a difference between variables: Larger, smaller, or about equal to 0. The most risky predictions a theory can make are point-value predictions (actual measurement outcomes), because the variation in possible outcomes of such future observations is countably or uncountably infinite. To test a prediction in the social and life sciences, most often a hybrid of Fisherian significance testing and the Neyman-Pearson paradigm of Null Hypothesis Significance Testing (NHST) is used (Nickerson, 2000 is a review of the use of NHST in social science).[11] Other fields may use different procedures, but generally speaking it involves estimating how likely it is to have observed the data, if the theory would not have urged us to make an observation (the chance of observing the data, or more extreme data, given that the or the null-hypothesis (H_0) is true). This concerns data patterns that could have revealed –*absent the theory*– the predicted observational constraint against the background noise of'obvious' observations. The predicted observation has to stand out above this background noise and the null-hypothesis test is a decision on whether to accept or reject the observation as being background noise. The difference between corroborating the prediction of a

[9]Admittedly, I had not heard of this word before I read'Sloughing Ontology' (Dale, 2008). Given our Theorrhea /Empiarrhea exchange, it seemed appropriate to use Epistemic Sloughing here.

[10]Strictly speaking the undirected Boolean prediction:'there is a difference / correlation' or'there is no difference / correlation', is weaker because it can just take on two values. The difference is lost due to the dichotomous nature of null hypothesis significance testing. The significance threshold is adjusted to represent the difference the two (one-tailed or two-tailed test).

[11]I will not consistently distinguish between the two approaches and refer to the whole enterprise of hypothesis testing as NHST.

positive correlation by rejecting H_0 and an expected instrument reading of 8 decimal points accuracy by rejecting H_0 is lost in the dichotomous accept-reject outcome (Cumming, 2012; Steiger, 2004). In the accept-reject dichotomy lies the meta-theoretical problem: Assigning a weight to the corroboration of the theory when H_0 is not refuted. The epistemic linkage between the substantive theory T, and the hypothesis test of its prediction **H***, has been broken (Meehl, 1997, 2002). The exoskeleton of truth-like scientific knowledge that supports the derivation chain of a theoretical prediction is effectively shed by hypothesis testing and interpreting the significance of a body of empirical evidence compares to picking scabs.

So why were scientists studying cosmology more impressed by GRT's convergence on predicting the orbit of Mercury? Meehl suggests it was the extreme numerical accuracy of the after-the-fact prediction. The bending of light phenomenon could also be derived from Newton's theory of gravity and GRT was indeed much more accurate than those predictions. This was however incomparable to the numerical accuracy of Mercury's predicted orbit. As an example of the odd consequences this epistemic sloughing can have on evaluating theories, consider the evidence for the existence of atoms and molecules. In 1913 Perrin made a famous argument in favor of the postulate that atoms needed to be included in the scientific description of reality by showing 13 qualitatively different ways to calculate Avogadro's number (amount of atoms in a volume of space). One could use the kinetic theory of gases, the electrochemical theory of electrolysis or X-rays to arrive at an estimate. Einstein's 1905 observations on Brownian motion were used and one derivation could be made based on the fact that the sky is predominantly blue during the day (Meehl, 1990a; Salmon, 1978). This argument convinced even the most persistent skeptics (except Ernst Mach apparently, see Meehl, 1990b), because it would be extremely unlikely if these different calculations would converge (by order of magnitude) close to $6.02214129(27)$×$10^{23} mol^{-1}$ based on chance alone (Meehl estimates the odds are over a quadrillion to one against). As Poincaré lucidly stated, if 13 ways to count something yield approximately the same number, something must have been counted (Meehl, 1990a).

The 13 estimates were however not exactly the same and this did not surprise any physicist, because the theories used in the derivation of the estimates (the auxiliary theories) were perceived to represent different degrees of verisimilitude. In an interesting twist, Meehl (1990b) explores what would have happened if in those days a statistical test were demanded to test the null-hypothesis'there is no difference between the estimates'. The result would be that H_0 should be rejected and as a consequence the atomic/molecular theory should be rejected. No physicist would have even contemplated declaring the theory refuted based on this evidence, because the epistemic relevance of the entire derivational chain of the argument leading to this convergence of estimates was uncanny. A concrete indicator of this relevance is the order of magnitude of the values (1023) that was recovered by all 13 estimations. The epistemic significance of this magnitude is lost if one considers the 13 estimates as a distribution of observations that should represent a true score, but take on different values due to measurement / estimation errors (T = X + E). To give an example of what physical science has achieved in terms of precision of point-value predictions, consider one of the most precise predictions of a measurement outcome by a scientific theory. It concerns the existence of an anomalous electron magnetic dipole moment (which is anomalous to the Dirac equation) by Quantum Electrodynamics (QED). According to the most accurate calculations the anomaly, when measured in a metal trap with a cylindrical cavity whose resonance structure is known, should have a value of: $a\mu^{QED} = 11659180.4(5.1) \times 10^{-10}$ (Aoyama, Hayakawa, Kinoshita, & Nio, 2008; Hagiwara, Martin, Nomura, & Teubner, 2007). The most accurate empirical measurement of the predicted value of the anomaly in the described measurement context is: $a^{exp} = 11659208.0(6.3)x10 - 10$ (Bennett et al., 2006). Indeed, the residual difference between theory and measurement is not zero, but the precision and accuracy of the theoretical prediction are unparalleled by any scientific theory (this residual difference can however be understood in Quantum Chromodynamics or QCD). No cats were killed in the process of measuring this accuracy

(Schrödinger, 1935).

Such examples of accuracy do not exist in any other discipline of science, moreover note that the theory in fact predicts the entire measurement context in which the phenomenon may be observed to yield the predicted measurement outcomes. It would be the same as having a psychological theory that yielded a formula that could be used to predict numerically the difference between running a test on a computer and filling out a paper and pencil version of the same test. In social science, the prediction of measurement context is implicit, but very real; consider the factorial design of experiments or the observation of a variable corroborating a theoretical entity only at the level of a sample of a certain size. In fact, I propose to consider all actions taken to make sure that'everything else being equal' the evidenced prediction should be attributed to the theory (the ceteris paribus clause) reflect expectations by the experimenter about the measurement context in which the phenomenon should be observable.

In principle, the statistical testing of hypotheses is not a problem as long as sensible scientists uphold the epistemic linkage between theory and tested hypothesis in their evaluation. The nature of the prediction made by the theory plays an extremely important role in this respect. The odds of a quadrillion to one against count as a'*damn strange coincidence*' (Salmon, 1984), but these odds are the exclusive result of point-value predictions. Therefore, in physics this will hardly present a problem (note, they often do not use statistical tests at all). In the softer empirical sciences, whose theories do not predict anything beyond the direction of an association between variables, a whole range of obfuscating factors make it extremely difficult for scientists who are genuinely willing to do so, to reconstruct the epistemic linkage in a meaningful way:

> `Thesis: Null hypothesis testing of correlational predictions from weak substantive theories in soft psychology is subject to the influence of ten obfuscating factors whose effects are usually (1) sizeable, (2) opposed, (3) variable, and (4) unknown. The net epistemic effect of these ten obfuscating influences is that the usual research literature review is well-nigh uninterpretable.' (Meehl, 1990c, p. 197)

If Meehl's thesis is correct and essential mechanisms for deciding the veracity of theoretical claims in a field of science are sabotaged such that its claims are uninterpretable, it cannot be considered a scientific enterprise that generates increasingly accurate knowledge about the structure of the domain in reality it studies. There may be other goals for genuine scientific endeavours to pursue, but I am considering empirical sciences with just cause: In this scenario we cannot expect the applied fields of the social and life sciences to produce effective and efficient technology (e.g., Ioannidis, 2005; Worrall, 2011). Table 1.3 lists Meehl's obfuscating factors of which many will eventually be discussed one way or another in this book. In this section I will address two categories of factors: 1) Factors that indicate a credibility hurdle is necessary for the empirical fields (1.3.1) and 2) Factors indicative of obscurantist practices concerning the derivation and evaluation of a prediction (1.3.2).

1.3.1 The crud factor and the credibility hurdle

Lykken (1968) estimated that the'unrelated' molar variables involved in most studies in psychology share 4-5% common variance, meaning, with 0 measurement error a correlation of about .20 can be expected between any one of them. This really depends on the field of inquiry, but it seems that estimates between .15 and .35 are by no means an exaggeration (Lykken, 1968; Meehl, 1990a, 1997). Based on the lower estimate, the expected difference between any group-based averages would be about 0.5 standard deviations. The test against the null hypothesis of'no association' is often a test against a'straw man' null hypothesis (LeBel & Peters, 2011), because it can be known in advance that an assumption of no association at all is false (Bakan, 1966; Bower, 1997; Ferguson & Heene, 2012; Gliner, Vaske, & Morgan, 2001; Meehl, 1967; Nunnally, 1960; Rozeboom, 1960).

Table **1.1**
Meehl's (1990a, 1990b, 1990c) Obfuscating Factors that render Scientific Claims Uninterpretable in the Context of NHST and Weak Predictions. The last Columns display the Effects of the Factors on the Perceived Verisimilitude of Theories that are in fact Truth-like and Trivial or Likely False.

	Effect on perceived verisimilitude of a theory that is in fact:	
Factor	*Truth-like*	*Trivial / False*
1 Loose (nondeductive) derivation chain, making several 'obvious' inferential steps requiring unstated premises (intuitive, common-sensical, or clinical experience).	-	
2 Problematic auxiliary theories, although explicitly stated.	-	
3 Problematic ceteris paribus clause.	-	
4 Imperfect realization of particulars (experimenter mistakes in manipulation) or experimenter bias in making or recording observations.	-	
5 Inadequate statistical power to detect real differences at the convent' ional significance level.	-	
6 Crud factor: In social science everything correlates with everything to some extent, due to complex and obscure causal influences.		+
7 Pilot studies used to (a) decide whether 'an effect exists' and (b) choose a sample size of adequate statistical power if the pilot effect is borderline but in the 'right' direction.		+
8 Selective bias in favor of submitting reports refuting the null hypothesis.		+
9 Selective bias by referees and editors in accepting papers refuting the null hypothesis.		+
10 Detached validation claim for psychometric instruments.	-	+

Therefore, a researcher can maximize his chances to corroborate any weak prediction of association between variables, by making sure a large enough number of data points are collected. This'crud factor' (cf. Meehl, 1990c) implies a researcher has a chance of 1 in 4 to evidence an association using a sample size of 100 data points, without even needing a truth-like theory to predict an association (Bakker, van Dijk, & Wicherts, 2012; Ioannidis, 2012; Simonsohn, Nelson, & Simmons, 2013). This is of course not something a researcher deliberately plans to do (one assumes), or a statistical error that is committed. It is a genuine problem of theory evaluation, rather specific for the domain of the social and life sciences. The problem does have a practical solution and this involves what may be called raising the credibility hurdle for observed phenomena.

By increasing the number of data points to evidence a pattern, one increases the statistical power of the study (its sensitivity) to detect an association when it is indeed truly present in the data as posited by the theory (a'true effect'). Suppose a researcher publishes a multi-experiment article that reports multiple corroborations of a predicted association between two or more variables. Why publish multiple experiments if the association was evidenced in the first experiment? One reason is purely meta-theoretical; to increase the credibility of the theory that predicted the phenomenon, add to its perceived verisimilitude (in empirical reports of psychological science, authors often report they attempt to show a certain effect is'real'). Assume H_0 was refuted in 5 independent replications at $p < .05$, this would be an impressive corroborative track-record for the theory that predicted the association. That is, if there ever were a truth-like theory that made a prediction. As explained above, this scenario could happen just as likely for any pair of'unrelated'

Table **1.2**
*The Power Failure (Decline of Total Power) for Subsequent Corroborations Using the same Sample Size. Values calculated for 3 Different Levels of Cohen's d, using G*Power 3*

N corroborations	Total Power	N for Large	N for Moderate	N for Small
1	80 - 81	52	128	788
2	71 - 72	52	128	788
5	0.5 - 0.9	52	128	788
10	0	52	128	788

variables, provided the sample sizes are large enough. Knowledge of such facts does not prevent the zealous researcher from the impression that the corroborations of his predictions are really due to the verisimilitude of his theorising (not to forget the ingenuity of his experimental methods of course).

It is possible to solve this problem by taking seriously the rules of statistical inference and the epistemic sloughing effect of the null-hypothesis test when corroborating weak theoretical claims. In the multi-experiment example, the probability of each subsequent rejection of H_0 being a false positive (rejecting H_0 when it is true, a type-I error) represented by the α-level, effectively reduces to .055 = .000000312 (Schimmack, 2012). This is approximately the'discovery level' used to evidence phenomena in particle physics, like the Higgs Boson (the 5σ criterion). At first sight, this drop in the probability of rejoicing over a false positive seems rather welcome, but what about the probability of committing type-II errors (false negatives)? The decrease of the effective α-level represents the fact that the distribution of values that could be observed if H_0 were true (the background noise) changes and becomes narrower. This means, everything else being equal, there is a drop in statistical power, sensitivity to detect the true effect with the same sample size.

To maintain equal power across subsequent observations, one needs to increase the sensitivity of the study for each observation that adds to the credibility of the phenomenon. In social science this often means increasing the sample size of the study, in physics one resorts to building a more sensitive measurement apparatus. Table **1.2** shows what happens to the Total Power (of the replications as a whole) if the sensitivity is not adjusted. At 5 subsequent corroborations the Total Power has already dropped below 1%. If H_0 is rejected at this sensitivity level, it is very unlikely the results represent a true effect. This is a very real problem, it was recently concluded by studying a large number of meta-analyses that individual studies in Neuroscience are severely underpowered ranging from 8% to 21% (Button et al., 2013).

Schimmack (2012) provides a table with requirements for a multiple-experiment study to maintain a sensitivity of 80% for each individual rejection of H_0 (Table **1.3** is an excerpt). In other words, in order to be credible as a streak of corroborative events of a theoretical prediction, the observational hurdle has to be increased. The observation of n+1 significant effects out of N attempts at the Total Power level, is unlikely to be due to chance alone (or the crud factor): The probability of observing 5 significant results in 5 studies whose Total Power is 50% is 0.0313. So in just 3 out of 100 five-experiment studies of the same Total Power, we would expect to see 5 significant results. That would probably qualify as a'*damn strange coincidence*' (Salmon, 1984) if it had occurred absent the theory predicting the observation. Schimmack (2012) calculates an incredibility index (IC-index) as the binomial probability of observing at least one non-significant result in the streak of corroboration events, given an estimate of the Total Power to detect the effect (post-hoc observed power). For this 5-study example, given a large effect size (Table **1.2**) the IC-index would simply be 1-Total Power = 96.9%. That's how incredible these results would be, given the sensitivity of the test. Should I have increased the observational hurdle as indicated by Table **1.3**, the IC-index would

Table **1.3**
*The Relation between Subsequent Corroborations Events by Statistical Significance and Sample Size needed to maintain 80% Power for each individual Corroboration. Values calculated for 3 different Levels of Correlation |r| (based on Cohen's d), using G*Power 3.*

N corroborations	Total Power Needed	Large (|r|=.4)	Moderate (|r|=.25)	Small (|r|=.1)
1	80	52	128	788
2	89.4	136	336	2068
5	95.6	440	1090	6750
10	97.8	1020	2560	15820

have been 4.4% and I would have been much more comfortable to proclaim to have evidenced an association predicted by a theory.

There is no tradition to raise the credibility hurdle for subsequent corroborations of associated variables in the social sciences and in combination with the crud factor a dangerous recipe for immobilizing the scientific method emerges. Suppose dyslexic and average readers in fact belong to the same population of normal variation of reading ability and the only'true' characteristic that separates these groups is some demarcation of ability in the lower end of the reading ability distribution (see Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992). Just by taking into account a crud factor, it may be expected that reading ability correlates with any other variable, meaning, if we sample a group based on low reading ability, this group can be expected to have low scores on any other variable as well. The structure of the networks in section 1.2 seems to be a logical result of this ambient correlation. To illustrate the consequences of the premise, consider a numerical example by Meehl (1990a; see also Meehl, 1990c). Imagine he is talking about reading ability (x) and some random other variable, for instance, amount of music education received during childhood (y):

> "I provide one simple numerical example to illustrate the point that a modest crud factor cannot be discounted in the metatheory of significance testing. [...], suppose that a representative value of the crud factor in a certain research domain were r = .30, not an implausible value from the examples given. We have a substantive theory T, and we are going to `test' that theory by a correlational study involving observable variables x and y, which, however, have no intrinsic logical connection with T and have been drawn randomly from our huge pot of observables. Assume both x and y are approximately normal in distribution. We dichotomize the independent variable x at its mean, classify each subject as high or low on the x trait, and compare their scores on the dependent variable y by a t test. With the mean standard score of the highs on x being .8 (at + 1 MD) and that of the lows being -.8, there is a difference of 1.6 sigma in their means. Hence the expected mean difference on the output variable is d = .48, about half a sigma. Assuming sample sizes for the highs and lows are around 37 (typical of research in the soft areas of psychology), we find that the probability of reaching the 5% level in a directional test is .66. So a theory that has negligible verisimilitude, and where there is no logical connection between the theory and the facts, has approximately a 2-to-1 chance of being corroborated provided that we were predicting the correct direction. If one assumes that the direction is completely chance (which in any real research context it would not be, for a variety of reasons), we still have a .33 probability of squeaking through with a significant result; that is, the empirical probability of getting a positive result for the theory is larger, by a factor of 6 or 7, than the .05 we have in our minds when we do a t test' (Meehl, 1990a, p. 125)

In this example the high/low cut is made at the mean of x, but if this were a study in developmental dyslexia and x were reading ability, the cut would be much more extreme (10th or 25th percentile). The probability of finding a positive result without a theory is likely somewhat higher

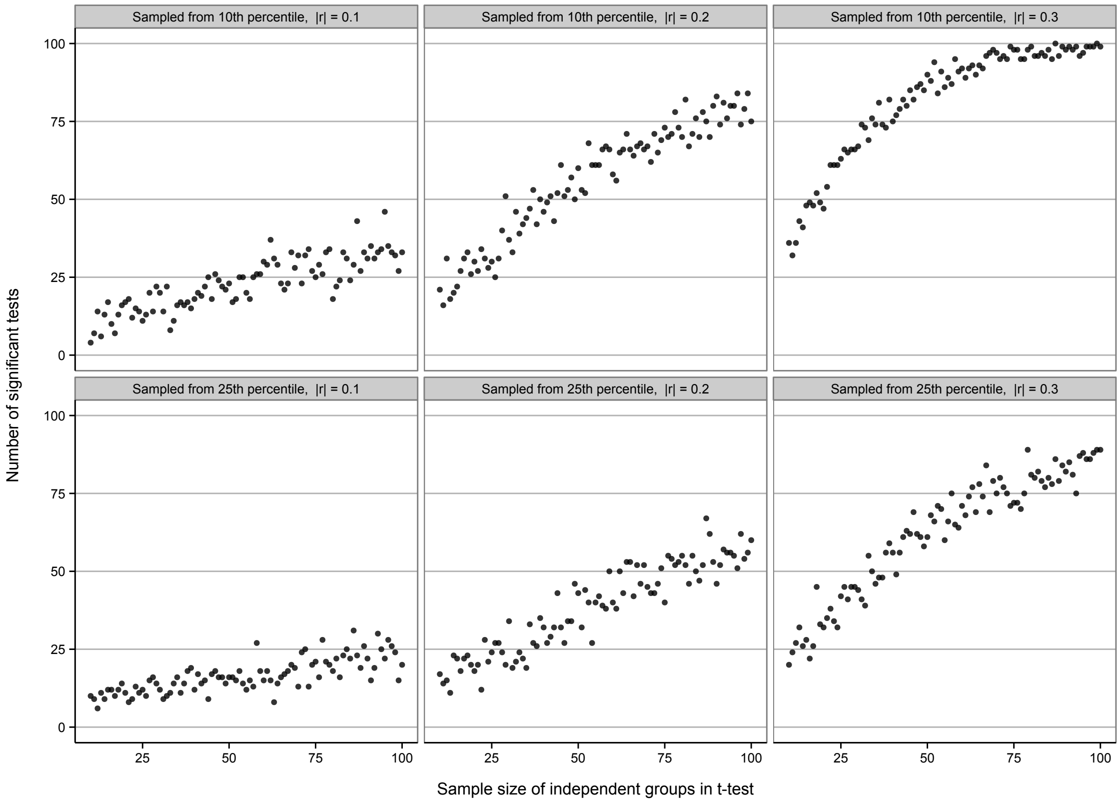


Figure 1.5 – A simulation of the effect of sampling from different regions of a population distribution (N_{pop} = 500000) in the presence of a crud factor, a population--level correlation between any two random variables. Each dot represents the number of significant results ($p < .05$) observed in 100 t--tests for independent groups of the size represented on the x--axis (10 – 100). Two random variables were generated for each population correlation: .1, .2, .3 (columns). One random variable was used to sample data points in the 10^{th} (top row) or 25th (bottom row), or between the 25^{th} and 75^{th} percentile (comparison group). The means concern the aggregated values of the second random variable for each sampled case. The directional hypothesis tested against the null was $(M_{[.25,.75]} - M_{[0,.10]}) > 0$ or $(M_{[.25,.75]} - M_{[0,.25]}) > 0$"

than 33% for a non-directional hypothesis, because in studies of developmental dyslexia one is always comparing whether the difference in means of performance measures is greater (or smaller) than 0.

Whether this complicates matters can be relatively easily tested. Figure **1.5** shows results that were obtained from a simulation of different levels of the crud factor and dyslexic reading selection criteria. First, three different population-level correlations values were used to simulate the crud factor association between any two random variables x and y (N_{pop} = 500000, with a correlation of x and y of .1, .2 and .3). Second, based on x (e.g., a standardised reading ability test), the population was cut into regions to sample dyslexic readers from (the 10^{th} percentile and lower, and the 25^{th} percentile and lower) and average readers (between the 25^{th} and 75^{th} percentile). These cut-offs were used to draw samples to conduct a t test for the difference between dyslexic and average readers in the sample on their mean value of y (e.g., amount of music training received). The sample size for each group was varied from 10 to 100 data points and 100 tests were performed for each group size. For each test a new random group sample was from the different regions of the population distribution.

The graphs represent the number of significant ($p < .05$) t tests found in the series of 100 tests conducted for each group size. If the correlation between any variable were .1, comparing to the samples from the 10^{th} and 25^{th} percentile would yield 25% significant results at group sizes of 44 and 58 data points, respectively. The total study sample size would be 88 and 116. At this crud factor level the chances do not get much better than 1 in 4 corroborative events without there being any theory to pat on the back and grant some verisimilitude. When the correlation is .2, 25% significant tests can be expected at group sizes of 12 (10^{th}) and 23 (25^{th}) and at a correlation of .3 it's 10 (10^{th}) and 12 (25^{th}) participants in each group to find 25% significant differences. The crud factor of .3 even implies that 100% of the conducted tests could give a significant result if group size is larger than 87 and the dyslexic group is drawn from the 10^{th} percentile of the population distribution of reading ability.

The neglected credibility hurdle for evidencing empirical phenomena raises an important point: Predictions of phenomena have to be risky (Popper, 1959), ideally they should be point-values, magnitudes, but interval predictions should also be possible (Cumming, 2012; Meehl, 1997). Predictions have to be severely tested (Mayo & Spanos, 2006), but perhaps the best characterisation that is relatively school-of-philosophy and discipline-of-science free, was given by Salmon (1984): A scientist has to make sure it would have to be a'*damn strange coincidence*' to have observed the predicted phenomenon, without guidance by the theory. Perhaps we should just acknowledge that in the presence of the crud factor, even if it is just .1, evidencing a significant difference between average and dyslexic readers on whichever performance measure you can imagine, is not that impressive. It does not qualify as a'*damn strange coincidence*' and therefore not as a corroborator of theories: Enter my characterisation of the state of the current scientific record on causes of developmental dyslexia.

1.3.2 Investigator James corroborates his muse: The EVA-ætiology

The crud factor is one of the many important factors listed in Table 1.3 that make the appraisal of theories in the context of weak predictions and NHST problematic, if not impossible. One more general cause underlying this notion are the unknowns, such as the number of pilot studies that were conducted prior to the publication of the corroboration of a prediction (factor 7, Table 1.3). This is in principle the same effect as just demonstrated in Figure 1.5, except the decision where to place the cut-offs is purely guided by a previous sample of values that turned out to yield significance of the hypothesis test.

There are other reasons than statistical ones, to be explicit about assumptions, expectations, unknowns and every decision that leads to a change of method or procedure (i.e., the opposite of being obscurantist). It is essential for understanding the deductive chain that led to a prediction, its logical structure (if any) and which link should be to blame, or be rewarded when the results are known. As a demonstration, I sketch an example of the evaluation of a novel ætiology for developmental dyslexia. As far as I know, this ætiology has not been proposed yet, but all the claims and derivations are based on the actual scientific record. To prevent any confusion about my actual scientific claims (e.g., due to out of context quoting), I will present the evaluation as a description of the theory of Investigator James, $\mathbf{T_{ij}}$[12]

Investigator James published a paper in which a prediction of the EVA-ætiology of developmental dyslexia, or, **Early Vestibular-Auditory deficit hypothesis** is tested. Recently a'*strong, early vestibular-auditory interaction that is critical for the development of music behaviour*' was evidenced in 7-month-old infants, who were able to encode rhythm in music using their body as evidenced by their bouncing behaviour. Encoding only took place if they could actively bounce by themselves while listening to the music. The authors suggest that'*the experience of body movement plays an important role in music rhythm perception*' (Phillips-Silver & Trainor, 2005). According to Investigator James this allows the derivation of a causal pathway for understanding developmental dyslexia, because:

[12]Investigator James is a zealous, but honest researcher who produces so many theories they have to be indexed T_{ij}. He is my terrestrial equivalent of Omniscient Jones, whose theories (TOJ) are always true (Sellars, 1956).

T_1: Rhythm, beat and meter perception in non-speech, speech and music is impaired in dyslexic readers as well as rhythm production (Corriveau & Goswami, 2009; Goswami, 2006, 2011; Huss et al., 2010; Overy, 2003; Ramus, Nespor, & Mehler, 2000; Thomson & Goswami, 2008) Results are accounted for by a theory proposing the real cause of dyslexia concerns timing deficits understood in a **neurobiologically** inspired **temporal sampling framework** (Goswami, 2011).

T_2: Balance and motor learning are impaired in dyslexic readers due to a **cerebellar deficit** (Brookes, Tinkler, Nicolson, & Fawcett, 2010; Ramus et al., 2003; Reynolds et al., 2003; Rochelle & Talcott, 2006; Stoodley et al., 2005; Stoodley & Stein, 2012). Exercise based training ameliorates language and literacy difficulties of dyslexic readers (Franceschini et al., 2013; Reynolds et al., 2003; Reynolds & Nicolson, 2007).

T_3: In average developing children and adults, a larger amount of lifetime music training is associated with better auditory discrimination, fine motor skills, a larger vocabulary, better non-verbal reasoning ability, better letter recognition and better speech in noise perception (Forgeard, Winner, Norton, & Schlaug, 2008; Proverbio, Manfredi, Zani, & Adorni, 2013; Strait, Parbery-Clark, O'Connell, & Kraus, 2013). A novel theoretical framework about **training-related plasticity** induced by the complex nature of learning, listening and playing music provides a **neuroscience explanation** for these associations (Herholz & Zatorre, 2012).

T_4: There is a **genetic factor** associated to **impaired speech perception and language production** (vocabulary) in infants of dyslexic parents, noticeable as early as 2-17 months of age (Been, van Leeuwen, & van Herten, 2008; Koster, Been, & Diepstra, 2005; Richardson, Leppänen, Leiwo, & Lyytinen, 2003; van Herten et al., 2008; van Leeuwen et al., 2006; van Leeuwen et al., 2008). These impairments in infancy are associated to reading and spelling problems later on in life (Guttorm, Leppänen, Hämäläinen, Eklund, & Lyytinen, 2011; Molfese, 2000). Speech perception based training ameliorates language and literacy difficulties of dyslexic readers (Tallal, 2004).

T_5: The genetic factor must be at work to impair auditory and speech perception in utero. It is known that there is **learning-induced plasticity of speech processing** before birth (Partanen et al., 2013) in addition to **speech perception, voice-pattern recognition and language learning** (DeCasper & Spence, 1986; Moon, Lagercrantz, & Kuhl, 2012). The cognitive and perceptual abilities that are higher in the population of well-trained musicians are typical of the abilities that are much lower in the dyslexic reader population. Given the conjectures of the musical training-related plasticity framework, it is likely that the temporal sampling theory and the cerebellar theory can be unified into a **general deficit in the fluency of coordination of perception and action**. Such a deficit was recently evidenced by examining long term correlations in naming latencies of dyslexic readers (Wijnants, Hasselman, Cox, Bosman, & Van Orden, 2012).

Using these 5 theoretical claims Investigator James believes he can predict an observational constraint between two unlikely variables that can finally resolve the long-standing debate on whether there is a speech perception deficit in developmental dyslexia (Ramus & Szenkovits, 2008; Serniclaes & Sprenger-Charolles, 2006):

Conjecture: *From* the theoretical claim about importance of early vestibular-auditory interaction and the experience of self-initiated bodily coordination to musical rhythm as crucial factors for the development of music behaviour and rhythm perception *it follows that* impaired music and speech rhythm perception together with impaired balance and motor learning skills in dyslexic readers are also due to the reduced in utero/neonatal/infancy learning-induced plasticity due to the general deficit of fluent coordination (**T_5**).

Prediction: *If* the amount of training-related plasticity induced by engaging in music related behaviour is associated with better letter recognition, a larger vocabulary and better auditory and speech perception in non-dyslexic readers *and if* action and speech training-induced plasticity can ameliorate reading ability of dyslexic readers *then* there must be a positive association between amount of music training received and speech perception performance in dyslexic readers (**T_1 + T_2 + T_3 + T_4 + T_5 + conjecture**).

Most of the theoretical entities used by Investigator James touch the data base, the empirical record of the field, but he posits some novel functional or compositional laws to exist between entities, like T_5 and the implication in the conjecture. The prediction is derived using at least five theoretical claims, each consisting of several auxiliary hypotheses and theories. Following Lakatos' terminology I will call the theoretical whole (**T_1 + T_2 + T_3 + T_4 + T_5**) the *hard-core* of the EVA-ætiology, or its *theoretical core* **T_{ij}**. Box 1.1 shows the elements involved in corroborating the theory. The sentence can be read as:'from **T_{ij}** *and* auxiliary theories *and* auxiliary instruments *and* assuming no interference of unknown factors and causes (*ceteris paribus clause*) *and* a truthful report of what transpired during the observation, *it follows that* if **O_1** is observed *then* **O_2** must also be observed.' The horseshoe is the symbol for material implication and the turnstile represents a deduction of the material implication based on the left hand side. **O_1** concerns evidencing developmental dyslexia and **O_2** concerns observing a positive correlation between amount of music training received and speech perception performance. The four possible theoretical outcomes for the observations are that **O_1** and/or **O_1** can be evidenced or not at all. However, the consequences for **T_{ij}** concern just two outcomes, being whether the material implication is true or false. By applying the valid logical syllogisms (see Table 1.5), it is clear there can be only one valid logical inference when appraising a scientific theory: If the right hand side is false, then **T_{ij}** is false (*modus tollens*). The other valid syllogism (*modus ponens*) is the derived prediction: **T_{ij}** is true and therefore the right hand side is true (the observational constraint). This is Investigator James' *assumption*, his *theoretical claim*, it is the *hypothesis* he wants to test. Only Omniscient Jones' (T_{OJ}) theories are true (Sellars, 1956) and therefore all a scientific community can do is decide how truth-like **T_{ij}** appears to them which means assessing its verisimilitude.

Let **O_1** and **O_2** be observed, meaning the right hand side of the corroboration formula is true. The only valid conclusion is that **T_{ij}** is not falsified. To claim to have evidenced the truth of **T_{ij}** is to commit the invalid third figure affirming the consequent. This is probably the most often committed invalid inference made in science and known as the'effect = structure fallacy'. It is often committed when inferring a cognitive modules exists based on each behavioural effect one observes (Bosman, Cox, Hasselman, & Wijnants, 2013; Van Orden & Kloos, 2003; Van Orden, Pennington, & Stone, 2001). Or, perhaps positing causal pathways for developmental dyslexia for each observed impaired performance measure evidenced in dyslexic readers as a group? Note that Investigator James, to derive the theoretical core of the EVA-ætiology, uses a mix of the syllogisms, mostly affirming the consequent, as if using modus ponens. This is based on a track record of just several corroborative events (the empirical studies cited), whose weight of corroboration we do not know much about. The ontological relevance of genetic and brain components does seep through in the weight attributed to some links in the derivative chain.

To be scientifically credible, or plausible, to enjoy a high perceived verisimilitude, the only thing **T_{ij}** can do is to make risky predictions that are severely tested. Not just once, it has to show a vita without any gaps, a track record of corroboration events (e.g., at least the 5 corroborative events at proper Total Power mentioned in the previous paragraph). More weight will be added to the corroboration if, absent **T_{ij}** the probability of observing **O_2** *conditional on* **O_1** is very low or is considered a'*damn strange coincidence*'. Known factors such as the crud factor and weak directional hypotheses should greatly diminish the corroborative weight awarded to observing the predicted observational constraint. In the case of **T_{ij}** one should ask: How extraordinary is the observation of a positive association between music education and speech perception in dyslexic readers, without **T_{ij}** suggesting the empirical inquiry? Perhaps there is some surprise left after discounting for the crud factor and the weak prediction, it is for the community to discuss. What would have guaranteed

Box 1.3: The corroboration formula for appraising a theory, adapted from (Meehl, 1990a; 1990b; 1997; 2002). The outcome in the case of a falsified conjunction (observation 2 was not evidenced) is given as two forms, one falsifying the conjunctions and its logical equivalent, falsifying at least one of its elements. Explanation of symbols: The dots state conjunctions ('and'), the $\vee$ disjunction ("or"), the turnstile ($\vdash$) signifies deductive derivability (entailment, 'from ...it follows that ...'), the horseshoe ($\supset$) is a material implication or conditional ('If ...then ...'), a tilde (~) represent negation ('not'), the three horizontal bars indicate full equivalence of the left and right hand side.

The corroboration formula for appraising a theory:

$$(T \cdot A_T \cdot C_p \cdot A_I \cdot C_n) \vdash (O_1 \supset O_2)$$

T : The theory of interest
A_T : Auxiliary theories relied on in the particular experiment
C_p : Ceteris paribus clause (other things being equal)
A_I : Instrumental auxiliaries (devices relied on for control and observation)
C_n : Realised particulars (conditions were as the experimenter reported)
$O_{1,2}$: Observations, or statistical summaries of observations

The case of the falsified conjunction $(O_1 \cdot \sim O_2)$:

$$\sim (T \cdot A_T \cdot C_p \cdot A_I \cdot C_n) \equiv \sim T \vee \sim A_T \vee \sim C_p \vee \sim A_I \vee \sim C_n$$

high corroboration strength without any discussion at all would be Investigator James' prediction of the exact value of the positive correlation, conditional on the severity of dyslexia of a participant (a point-value prediction). Unfortunately, that is beyond the scope of current theory.

1.3.3 When is a Lakatosian defence defendable?

What happens if the right hand side is false (the case of the falsified conjunction in Box **1.3**)? A consequence of the logical structure of the formula is that the falsification by modus tollens affects all the elements in the left hand conjunction that give rise to the derived prediction of observations. This means that the'cause' for not observing the predicted observations, the falsification, can apply to each (at least one) individual element in the formula. This can be called a'Lakatosian defence' as it represents the important amendments Lakatos added to Popper's logic of scientific discovery (Lakatos, 1970, 1974; Meehl, 1990b; Popper, 1959). Oversimplifying, in Popper's version of science, there would only be a T and modus tollens would dictate scientist to abandon the theory in an act of'instant rationality' (Meehl, 1990a). As in the case of Avogadro's number, molecular theory was not abandoned because the estimates weren't exactly the same. Some of the auxiliaries did not enjoy as much perceived verisimilitude as others and the community knew this. What eventually happened was that auxiliary theories used in the derivation chain were amended or replaced by more accurate versions which give the very accurate number presented earlier.

A Lakatosian defence allows a community of scientists to examine what in the derivation of the prediction might have caused the failed observation, without immediately abandoning the theory. It is possible a theory may need to be amended or that experimental procedures or instruments

Table **1.4**
The Four Figures of Deductive Inference. Two Syllogisms Constitute Valid Logical Inferences.

Figure	Statement	Deduction	Name	Conclusion
1	if p then q	$p \therefore q$	Modus ponens ("establishing mode")	Valid
2	if p then q	$\sim p \therefore \sim q$	Denying the antecedent	Invalid
3	if p then q	$q \therefore p$	Affirming the consequent ("effect = structure fallacy")	Invalid
4	if p then q	$\sim p \therefore \sim p$	Modus tollens ("destroying mode")	Valid

need to be refined. In my opinion this is a sensible thing to do in a science that has a tradition of seeking rigorous tests of theories that would maximise the corroborative strength of an observation, like a program of *strong inference* (Platt, 1964). Crucial experiments devised to test divergent predictions of competing theoretical claims. What is needed is a science that is able to keep track of the verisimilitude of the theories it produces in a more or less formal way (use meta-theoretical tools). Or, it has to be relatively clear to all practitioners of the science in question whether a theoretical prediction is supported by the data or not and which finite set of testable factors could cast doubts on that assertion (i.e., physics). This ensures that eventually the options for deflecting a modus tollens falsification death ray of instant realism to other factors besides T will be depleted in a relatively short period of time. In a science that does not formally and rigorously appraise its theories a situation could arise that should be somewhat familiar to the reader by now: Many different theories, apparently corroborated by empirical evidence, but the evidence has no authority to discorroborate or negatively affect the perceived truth-likeness of other theoretical claims. Returning to Investigator James' theory, let the observational conjunction be falsified, the positive correlation was not observed in dyslexic readers. What are the credible options for saving $\mathbf{T_{ij}}$ from refutation?

What is often observed in empirical reports of the social sciences is that causes for a (partial) failure to corroborate a prediction are attributable to unknown or unexpected factors that may have influenced a result. This means the ceteris paribus clause (**Cp**) is sacrificed in order to save **T**. **Cp** states that under the conditions of the test (e.g., the experiment), there are no causes other than those in the derivation chain that could lead to the predicted observation. Assuming **Cp** is true prior to the experiment is extremely important for the corroboration of the derivational chain. Assuming **Cp** must have been false after a failed test is the easiest way out of falsification of **T.** Common statement are:'sample was too heterogeneous','effect was observed after including an [ad hoc] moderator'. In the field of developmental dyslexia one can find discussion sections in which authors ask whether the dyslexics were'real' dyslexics, whether IQ was taken into account, biological age vs. academic or reading age, comorbidity. Chapter 2 and subsequent chapters provide a possible explanation for the easy and apparently credible Cp sacrifices and suggest context relativity of measurements should be a part of the derivative chain of observations. It has been known for some time that the tendency to make sure that'all other things being equal' is true, by imposing strict procedures for sampling, selecting and matching of participants, and so forth, basically means that verisimilitude will apply to that context alone. For studying properties of particles in a collider, that may be fine, but when a medical procedure has to be introduced into the public domain, it often turns out that theoretical causes (a pharmaceutical substance) proven very effective in a randomised control trial (RCT), turns out to be much less effective (= efficient) when introduced into society (Cochrane, 1972; Higgins, Green, & Collaboration, 2008).

The $\mathbf{C_n}$ part of the equation has received a lot of recent attention in psychological science, some of which were be briefly discussed in the preface (e.g., questionable research practices, p-hacking). The essential point is, if one cannot trust authors to provide full disclosure about everything one needs to know about the particulars of the predicted observation, its verisimilitude cannot be assessed (a recent call for public disclosure of particulars revealed some authors may violate this clause without even knowing it, see Lebel et al., 2013).

For $\mathbf{T_{ij}}$ and the EVA-ætiology the degree of truth-likeness of the theories that have been used to derive the prediction are important. These auxiliary theories and hypotheses ($\mathbf{A_T}$) have been used to infer 'modus ponens'-like relations connecting different theoretical entities (e.g., functional or compositional laws), without actually appraising them. In addition, the verisimilitude of the entities is unlikely to be known in any formal way, as in a track record of known precision and accuracy of corroborative events. In the case of $\mathbf{T_{ij}}$ the theoretical core consists of five theoretical statements (i = 1, … 5) and the conjecture'binds' them in a 'modus ponens'-like manner in order to allow the actual prediction of the observational conjunction. Within each of the five postulates that constitute the core, there are many auxiliary hypotheses and theories (indexed by j = 1, … n). I estimate on average 5 theoretical claims or events that corroborate a hypothesis within each core postulate are important enough to be blamed for not observing what was predicted. Ranging from the inference of a genetic factor based on infant prospective studies of children at risk for dyslexia to the increased speech in noise perception in non-dyslexic readers due to music training. There are at least 25 possible targets to which refutation may be deflected before Investigator James' actual prediction or the theoretical core is affected. Lakatos called this the protective belt, preventing the core from falsification. If ad hockery is allowed in a science, the protective belt of sacrificial auxiliary hypotheses is of course infinite. In the spirit of this book, I will use protective boundary to denote this phenomenon and proclaim that tests of predictions and appraisal of theories, should aim *beyond the protective boundary*, straight for the core. A sobering note about this objective, technically, the reports of corroborative events that constitute a substantial part of the 25 possible auxiliary sacrifices, will have a similar corroboration formula associated with the observation, the auxiliary takes on the role of **T** but with its own personal stock of sacrificial of auxiliaries. It almost seems as if there were a dark plot to keep us all occupied indefinitely.

Finally, one might suggest the use of meta-analysis as a tool to gain some confidence about the credibility of an entity or law. These analyses report a summary effect size of a particular predicted phenomenon, and magnitudes of effect sizes are generally not predicted by theories, mostly signs of associations. Using a meta-analysis for these purposes (to gain credibility for an effect) is committing the effect = structure fallacy quite literally, because no effect magnitude has been predicted except for a lower bound, the critical value associated with the α level. The summary effect sizes can be interesting to asses after the fact, to see if a study included in the meta-analysis was sensitive enough to have detected the summary effect size (Button et al., 2013). As a corroboration of predictions meta-analysis does not yield any empirical accuracy or precision outcomes that should comprise the track record of the theory. From the meta-theoretical perspective, meta-analysis can help to improve precision and sensitivity of future measurements by quantifying the sources of variation between studies that are supposed to measure the same phenomenon.

Meanwhile, Investigator James' eagerly awaits the reviews of his new and improved funding proposal to further test the EVA-ætiology, using RCTs this time. A systematic review of 851 studies on the relation between music education and language and literacy development of dyslexic readers had of course nothing to do with his design change and new focus of inquiry:

> "There is no evidence available from randomized controlled trials on which to base a judgment about the effectiveness of music education for the improvement of reading skills in children and adolescents with dyslexia. This uncertainty warrants further research via randomized controlled trials, involving a interdisciplinary team: musicians, hearing and speech therapists, psychologists, and physicians.' (Cogo-Moreira et al., 2012)

1.4 The tenuous nomological net: Theories of construction vs. theories of principles

The problems with theory evaluation presented so far are not new and can be appended to a long list of critiques tracing back to the earliest conceptions of some fields of scientific inquiry. For example, Ladd (1892) reviewing William James' 'The Principles of Psychology' (1890) concluded that establishing psychology as a natural science was an'utter impossibility'. James' had suggested that psychology was already using the methods of the natural sciences to test deep hypotheses about its object of study, while in according to Ladd:

> "[...] psychology as a science, devoid of all postulating of `deeper-lying entities,' does nothing of the kind. It assumes only the phenomena - the thoughts and feelings as actually known, and the possibility of ascertaining uniform relations among them.' (Ladd, 1892, pp. 29–30, emphasis added)

Based on some of the work discussed in the previous paragraphs, Lakatos classified the kind of theorising practiced in the empirical social sciences as one of the worst kinds of ad hockery{footnoteThere is also honest ad hockery: "In our theories, we rightly search for unification, but real life is both complicated and short, and we make no mockery of honest ad hockery" (Good, 1965) and predicted grave consequences if it were allowed to continue:

> ``After reading Meehl [1967] and Lykken [1968] one wonders whether the function of statistical techniques in the social sciences is not primarily to provide a machinery for producing phoney corroborations and thereby a semblance of 'scientific progress' where in fact, there is nothing but an increase in pseudo-intellectual garbage. [...] Or, as Lykken put it: 'Statistical significance [in psychology] is perhaps the least important attribute of a good experiment; *it is never a sufficient condition for claiming that a theory has been usefully corroborated*, that a meaningful empirical fact has been established, or that an experimental report ought to be published.' [...] Thus the methodology of research programmes might help us in *devising laws for stemming this intellectual pollution which may destroy our cultural environment even earlier than industrial and traffic pollution destroys our physical environment.*" (Lakatos, 1975, p. 176, footnote 1, emphasis added)

Why do the softer fields of science rely on NHST, or more generally speaking, significance testing of directional predictions to'ascertain facts', when it has been pointed out (by the very people who invented the techniques!) that to do so is logically flawed, mostly trivial and often just plain wrong? (Bakan, 1966; Carver, 1993; Cohen, 1994; Hogben, 1956; Lykken, 1968; Mayo & Spanos, 2006; Meehl, 1967; Michell, 2009; Neyman, 1969; Nix & Barnette, 1998; Nunnally, 1960; Ring, 1967; Rosnow & Rosenthal, 1989; Rozeboom, 1960; Steiger, 2004; Trafimow, 2003; Tukey, 1960a, 1960b; Wilkinson, 1999). Why doesn't it look for better formal tools, or just hire a mathematician to create them? In fact, it was sir R.A. Fisher himself who explicitly warned about the dangers of adopting the abstract mathematical concepts of the theory of probability and measurement error, without carefully examining whether they are appropriate for social science. That is, in service of the goals of improving'natural knowledge' about relevant phenomena, as is customary in the natural sciences:

> ``I am quite sure it is only personal contact with the business of the improvement of natural knowledge in the natural sciences that is capable to keep straight the thought of mathematically-minded people who have to grope their way through the complex entanglements of error, [...] Certainly there is grave confusion of thought. We are quite in danger of sending highly trained and highly intelligent young men out into the world with tables of erroneous numbers under their arms, and with a dense fog in the place where their brains ought to be." (Fisher, 1958, p. 274; also see Yates, 1968, who reiterates Fisher's point a decade later).

In the sections that follow I will attempt to explicate Fisher's'grave confusion of thought' about'the improvement of natural knowledge' in the soft sciences. These observations warrant a broader discussion of problems with theory construction in the soft sciences before suggesting novel directions for the scientific inquiry into the ætiology of developmental dyslexia. As I referred to in the preface, we need to respect our elders (and sometimes their frustration and discontent and perhaps their failures as well), because I see history repeating itself and this time there is even a tendency to feign ignorance about the severity of non-replicability of phenomena, hypothesising after the results are known (HARKing; Kerr, 1998) underpowered studies, p-hacking and data-peeking to decide on sample sizes. To claim ignorance on those issues is to deny the historical scientific record. I did not cite some obscure scholars just to taunt scientist of the softer fields, of which I am one myself[13]. and although encouraging practices like data sharing or conducting confirmatory and replication studies is very necessary, it will not suffice to save the empirical social sciences from itself. One must ask, what is so special about the way the natural sciences theorise about the constituents of reality?

1.4.1 Constructive and Principle Theories: Synthesis versus Analysis

At the time of the 'confirmation' of some of the testable predictions of the theory of General Relativity discussed in paragraph 1.3, Einstein wrote a lucid letter to the London Times (November 28, 1919) in which he characterised two kinds of scientific theory:

> ``We can distinguish various kinds of theories in physics. Most of them are constructive. They attempt to build up a picture of the more complex phenomena out of the materials of a relatively simple formal scheme from which they start out. Thus the kinetic theory of gases seeks to reduce mechanical, thermal, and diffusional processes to movements of molecules – i.e., to build them up out of the hypothesis of molecular motion. When we say that we have succeeded in understanding a group of natural processes, we invariably mean that a constructive theory has been found which covers the processes in question.
>
> Along with this most important class of theories there exists a second, which I will call ``principle-theories.'' These employ the analytic, not the synthetic, method. The elements which form their basis and starting-point are not hypothetically constructed but empirically discovered ones, general characteristics of natural processes, principles that give rise to mathematically formulated criteria which the separate processes or the theoretical representations of them have to satisfy. Thus the science of thermodynamics seeks by analytical means to deduce necessary conditions, which separate events have to satisfy, from the universally experienced fact that perpetual motion is impossible.
>
> The advantages of the constructive theory are completeness, adaptability, and clearness, those of the principle theory are logical perfection and security of the foundations.''

-Einstein (1934/1952)

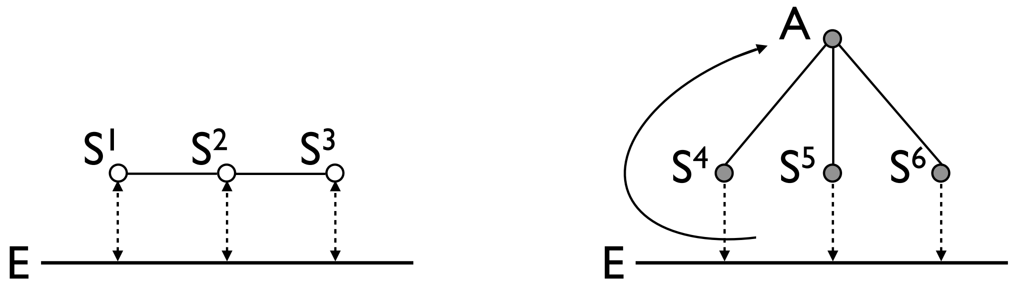
Contemporary physical theories are mostly principle theories and although he described constructive theories as "most important" it is evident that Einstein believed only principle theories could advance fundamental scientific knowledge about the universe. To explain the profound difference between the two types of scientific theory, the metaphor of the nomological net is again helpful. A part of the nomological net representing a scientific description of a domain in reality that is logically strong with a track record of strong corroboration and thus perceived as having high verisimilitude, may look something like the right panel of Figure **1.6**. This is an adaptation of a drawing by Einstein used to explain his views on scientific theorising in a letter to his friend Maurice Solovine (see Figure 1.6a)[14]. This would be a rather coarse rendition of a nomological net as

[13] This appears to be a sentiment among some peers:'replication bullies' and'data detectives' ruin their careers by pointing out statistical errors in their published work. (See e.g., Schönbrodt, 'About Replication Bullies and Scientific Progress... Retrieved May 2014 from: http://www.nicebread.de/about-replication-bullies-and-scientific-progress/)

[14] I thank Michael Seevinck for introducing me to this important history of science.

Theory of construction
a maturing science

Theory of principles
consensus formalism science



- *"devoid of all postulating of 'deeper-lying entities' "*
  There is no (explicit) formal system of axioms and postulates. A theory (**S1-3)** is constructed by association or conceptual replication. There is no consensus in the scientific community about the system it studies, what the interesting phenomena and levels of analysis are.
- *"It assumes only the phenomena"*
  Phenomena are **induced** directly from the manifold of immediate sensory experiences **E**, as empirical laws **S**.
- ***"****possibility of ascertaining uniform relations among them"*
  Predictions by a core of **S** concern the observation of correlated phenomena at some level of confidence. **E** informs and adjusts **S** by induction.
- *Theory Evaluation*
  Step I: Identify the theoretical core (e.g. **S1**, **S2**, **S3**)
  Step II: Identify the core hypotheses
  Step III: Scrutinize the empirical evidence
  Step IV: Estimate predictive power and empirical accuracy

- *Consensus Formalism*
  A formal system of axioms **A**, is raised from conjecture to postulate (the upward arrow). **A** describes a specific domain in reality for a discipline of science to study; a system, its relevant phenomena and levels of analysis. There is a consensus in the scientific community about the truth status of the formalism.
- *Closed Theory*
  Laws **S** are **deduced** from the formalism and a theory (**S4-6**) competes for scientific credibility within the domain described by **A**. There is no logical route from **E** to **A**, only from **A** to **S** to **E**.
- *Accurate Prediction*
  A theory predicts a measurement context in **E** in which phenomena may be observed in terms of measurement outcomes. **E** cannot adjust **S**, a prediction is accurate or not.
- *Theory Evaluation*
  Competing predictions are tested under conditions of strong inference; predictive power and empirical accuracy are assessed over a number of such tests.

Figure 1.6 – Adaptation of the "Solovine Schema (see van Dongen 2010, p. 52-53). The quoted text is from Ladd's 1892 characterisation of James' vision of psychology as a natural science.

the nodes **A** ('System der Axiome' [system of axioms])[15] and **S** ('Gefolgerte Sätze' [deduced laws]) represent collections of many different kinds of theoretical entities. As explained earlier, in terms of the nomological net metaphor (cf., Cronbach & Meehl, 1955), the strands of the net are supposed to represent the functional or constitutive laws that connect theoretical entities. Turning to a more stylised version in Figure **1.6** (right column), it is important to note that from **A**, lawful relations between theoretical entities **S** can be deduced and those lawful relations predict observable phenomena in **E** ("Mannigfaltigkeit der unmittelbaren (Sinnes-) Erlebnisse' [manifold of immediate / direct sensory experiences]). One or more **S** together can be regarded as a theoretical whole, a scientific theory (e.g., $\mathbf{S_{4\text{-}6}}$), the theoretical core.

What makes these theories principled theories? Perhaps surprisingly, it is not the logical strength of a pathway from observations of phenomena in **E** to a system of axioms **A** formally describing a domain in **E**. According to Einstein such a logical route from **E** to **A** does not exist at all. The process of theory construction always starts with a creative-intuitive act in which a scientist raises a conjecture to a postulate (cf., van Dongen, 2010). This is the upward arrow from **E** to **A** (the system of axioms) and it represents the bold, but inspired act of a scientist laying explanatory claim to a certain domain in reality, by stating:'let's see what happens if we assume it really is like this'. As mentioned earlier, in the hard sciences it is often the case the definition of a formalism is a community effort and therefore there is consensus about its verisimilitude. The purpose of a scientific theory departing from within the domain defined by the axioms and postulates of the formalism is to predict observable phenomena in **E** with a higher precision and accuracy than predictions by competing theories that depart from the very same formalism.

[15] I will refer to the system of axioms as the formalism hereafter. A full mathematical axiomisation of a physical theory is rare (Bunge, 1967).

The actual test of the precision and accuracy of the predicted phenomena constitutes an appraisal of the verisimilitude of the theory whose outcome could ultimately have consequences for the perceived truth-likeness of the entire formalism as well. Theories based on the same principles allow for a program of strong inference (e.g., Platt, 1964): A single datum can be obtained in a crucial experiment to decide between the veracity of competing theoretical claims. The logical strength of the derivational chain leading to the prediction is such that '*It is the theory that decides what may be observed*' (Einstein as quoted by Heisenberg, 1971, pp. 62–63). The single headed arrows represent explanation by D-N (Table 1.2) and as such, principled theories can be characterised as 'closed' (cf. Bokulich, 2004). They break down as a whole when a crucial experiment does not yield the results predicted by the theory: '[...] *if an experiment does not fit in Newtonian physics, you don't know what you mean by the words*.' (Heisenberg interviewed by Kuhn, 1963, p. 24 , February 27th).

> ``The chief attraction of the theory lies in its logical completeness. If a single one of the conclusions drawn from it proves wrong, it must be given up; to modify it without destroying the whole structure seems to be impossible." (Einstein, 1934/1952)

Of course, for theories of such logical strength that can be conclusively be shown to be false, a Lakatosian defence is permitted if the circumstances allow it.

In theories of construction, depicted in the left panel of Figure 1.6b, a theoretical whole (e.g., $S_{1\text{-}3}$) is 'hypothetically constructed' instead of 'empirically discovered' by proposing associations between phenomena described in the empirical record exist. In the most exemplary case there are no restrictions whatsoever on forming associations between observed phenomena (Einstein called this 'adaptability' in the quote form the letter). The left panel of Figure 1.6 represents a tenuous nomological net of theoretical entities that are all connected to, and therefore defined by empirical observations (exclusively I-S or sometimes D-S explanations, hence the double-headed arrows). There is no (explicit) abstract formalism defining a domain in reality that could prevent theoretical entities to become associated with one another in order to constitute a theoretical whole. This is why it cannot be said they are empirically discovered by testing risky predictions, their adaptability can make them 'complete': Any theoretical entity can become associated to any other entity by constructive hypotheses. To identify a theoretical core, one will often look for the least disputed empirical results and the hypotheses that predicted them. The options to avoid refutation, as explained in section 1.3, will be virtually limitless if the epistemic link is severed. The protective boundary will specifically serve to save those phenomena that support the theoretical core.

This bears strong resemblance to the situation described for theoretical accounts of the ætiology of developmental dyslexia in the previous paragraphs: The node categorising theory words was associated to all the other nodes that mainly captured different experimental designs or performance measures, hence empirical phenomena (e.g., brain, attention, learning, auditory perception). As a consequence, all the theoretical entities and laws that constitute the theory are very direct descriptions of empirical phenomena observable as a manifold of direct (sensory) experiences (e.g., "*it assumes only the phenomena*"). That is, in most cases, to describe theoretical entities in a theory of construction one can suffice with a level of abstraction that does not rise above a common language description: speech-sound perception, 'rules for converting mental representations of letters into sounds', 'a part of the visual system capable of fast signal transmissions'. Of course, in the literature these entities be referred to by their own proper neologism, but this is very different from the level of abstraction necessary to describe observables corresponding to non-commutative operators in a quantum measurement or a Ricci tensor in relativistic cosmology (both are mathematical abstractions that are hard to relate to any perceptual experiences).

The derivational chain leading to a novel prediction in a theory of construction is weak, or soft and so are the topologies of its nomological nets, even so, "*he who casts, will catch*". In most cases the chain of 'derivations' is based on previously observed particular facts (e.g., "[...] ascertaining uniform relations among [phenomena]") associated by probabilistic laws (see Box **1.2**). As a consequence, the failure to observe a predicted phenomenon does not nearly have as many implications for the perceived veracity of the theoretical entities involved, as is the case in a theory of principles. The scientist who uses theories of construction to study the structure of reality refrains from "postulating deeper-lying entities" (Fanelli, 2010) and constructs and amends a theoretical whole guided by phenomena directly observable in **E**. In a play of words with Einstein's dictum, in the softer sciences: "*It's the observations that decide what may be regarded as the theory*".

1.5 The thin ontic line (and why it must be cut)

1.5.1 This is the story so far:

In section 1.2 the impression that a plethora of different theoretical claims about the causes of developmental dyslexia are considered to be'true' at the same point in time or at least the impression that they escape falsification by means of discorroborative empirical evidence, was made plausible by a quasi-historical, quasi-nomological network analysis of terms used in abstracts of scientific articles on the subject.

In section 1.3 meta-theoretical causes were proposed to explain the erosion of the corroborative authority of empirical evidence. Important causes for the field of dyslexia research were identified such as the crud factor, epistemic sloughing by hypothesis testing and the inappropriate use of the Lakatosian defence.

Section 1.4 provided a meta-theoretical taxonomy of theorising in science and identified the theories produced by the social and life sciences as theories of construction that fail to achieve a level of abstraction of reality beyond natural language descriptions. They refrain from positing phenomena at deeper-lying levels of reality.

Which identified *cause* is sustained by which identified *effect*, or vice versa, is difficult to tell and most likely futile to attempt to figure out. One half of the arguments borrow authority from history, the other half from philosophy and logic, and sprinkled on top are some extra crunchy authorisations from physics. But what does this strange concoction authorise? To me, it authorises the rest of this book, including its slightly ridiculous late coming of age. It authorises a critical and more formal examination of scientific theorising when applied to inquire about the developmental origins of an observable state of impaired performance in young children and of the technology developed to alter that state. First, I sketch the framework for theorising and theory appraisal in the softer fields of science that emerges from the story so far.

1.5.2 Theory evaluation in the empirical Social and Life Sciences [a summary]

T_{ij} is posited by a scientist in the softer fields of science as a set of connected statements about entities and laws. The text in which $\mathbf{T_{ij}}$ is embedded is not an operational text. That is, there is no separate formalism containing theory language defining the entities and laws that an operational text commonly connects to observational language. There is no formal calculus or'truth-grinding machine' that can be used to decide on logical structure of derivations and verisimilitude of corroborations. The embedding text may be something akin to an interpretative text. This is a text in which theoretical concepts are defined in terms of the theory language alone. Theories of construction contain mostly observation language, so their embedding texts, that is, the scientific publications in peer-reviewed journals of soft empirical science, define operational theoretical entities in terms of other operational theoretical entities.

A soft embedding text can still be used by a meta-theorist to extract a lot of information from the set of statements. The text may contain some hidden theory language and a large subset of laws and entities will be associated to empirical evidence by observational predicates and statements. This often means formal language is introduced in corroboration (measures of association, functional forms, etc.).

T_{ij} contains postulates that can be categorised as central (i) or peripheral (j). There are just a few central, or *core postulates* and associated to each of them are many peripheral, or *auxiliary postulates* ($\mathbf{A_T}$) that have to have some degree of truth-likeness. *Auxiliary instrumentation* ($\mathbf{A_i}$) the theory relies on is also a part of the periphery. In the case of corroboration by prediction of empirical observations, a ceteris paribus clause ($\mathbf{C_p}$) has to be adopted together with an assumption about the truth of statements about realised particulars ($\mathbf{C_n}$). The truth status of the latter clauses represents the basic conditions for an empirical science to work. Improving assertions about their truth is the topic of many contemporary discussions in social science.

T_{ij} uses entities in derivational chains: Mental representations, genetic and neural components (structures, complex), developmental milestones (events), persistent disabilities (states), and realisable dispositions such as training induced plasticity to ameliorate language and literacy. Entities may be classified by a meta-ontology that I will refer to as mentology[16]. An example is *Meehl's ontology* (Meehl, 1993). According to this mentological list that has been useful for appraising psychological theories, the world consists of substances, structures, events, states, dispositions and fields. In softer fields, structures are almost always complexes, never simplexes like quarks. As different parts of a complex can have different verisimilitude, appraisal of such theoretical entities is difficult without formal definitions of the components of a complex.

T_{ij} uses laws to connect entities in derivational chains: A mental representation of speech sounds includes frequency and amplitude information (compositional / structural law, defining the parts and their arrangement). A mental representation that is not composed according to a law can cause a state of impaired reading ability (functional / dynamic law, change processes). In mentology there are also developmental laws, they represent a combination of the compositional and dynamic laws often pertaining to variables that are not under control, such as in the'documentary disciplines": history, palaeontology, geology and evolution. Darwin's theory contains a developmental law.

Laws are the statistical or formal regularities that exist between individual entities: The sign of 1^{st} and 2^{nd} order derivatives, partial derivatives (order of signs, factor interaction), functional form (exponential growth function, predator-prey dynamics), rank order of parameter values, or actual parameter values (point-value prediction) of functional relations.

T_{ij} may be represented as a nomological net: Entities constitute the nodes, the laws the edges. A proper subset of the net makes contact with reality in that it is an observational subset. Some entities can be directly related to observations by I-S inference without violating Einstein's dictum, or the logicians' corroboration formula. To logically *re-construct* theoretical propositions such that they acquire empirical content can be done using the so-called Ramsey Sentence. Einstein posited the creative intuitive act to cut loose from reality into abstraction, but'something' empirical is always retained. This implicit phenomenon was named'Ramsified upward seepage' (Meehl, 1990; Meehl, 1978, 2002, 2004b). What is important to note is that according to meta-theory it is possible to assert and define an entity at the same time, not in the operationalist sense, but by using a system of formal expressions.

[16]Mentology could represent a contraction of Meehl's Ontology or Meta / Mental Ontology, but more importantly it represents the fact stressed by Meehl that any attempt at a general classification of the lynchpins of reality depends on the mental lynchpins of the scientist. Meehl frequently solicited suggestions for change, expansion or reduction.

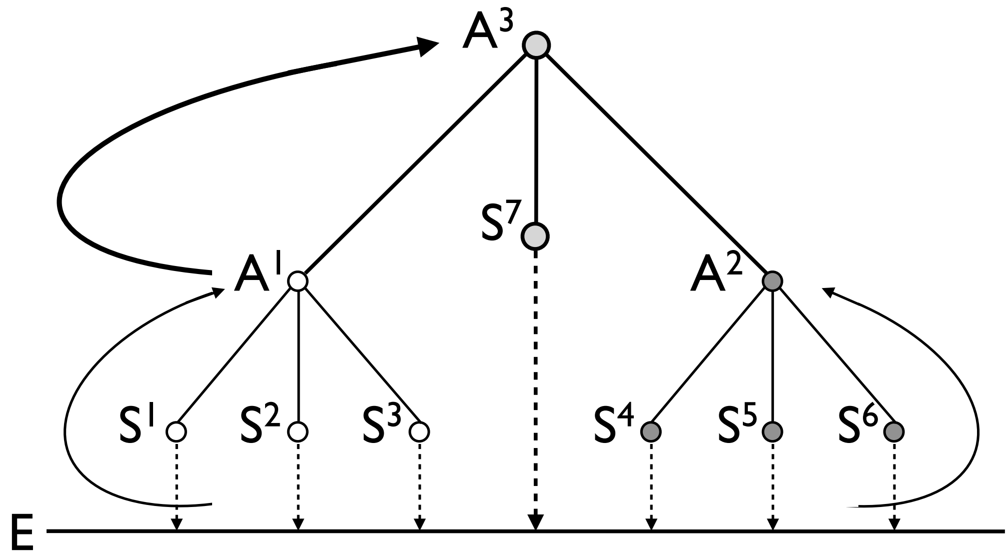


Figure 1.7 – Einstein's unification (see van Dongen, 2010) and the consequences in terms of explanatory power, predictive power and empirical accuracy of theories by increasing distance from the manifold of direct experience, **E**.

Falsification of a theoretical core in soft science does not occur very often. The same holds for corroboration or discorroboration by replication of observations, because such studies are rarely conducted until very recently (Klein et al., 2014). Much of the perceived verisimilitude of core theories in soft science is based on corroborative singularities. Black Hole phenomena in the nomological net that can only be approached as far as the event horizon. Their structure remains obscured to the inquisitive scientist, no matter how much daylight is let in to illuminate it. More formally stated, a Lakatosian defence is used to deflect falsification of the core to auxiliaries, the periphery. Or, the truth status of the ceteris paribus clause is sacrificed and an ad hoc auxiliary hypothesis is postulated. If an identified $\mathbf{C_p}$ falsifier can be changed into a corroborator on subsequent inquiries, this is a *progressive research program*, because a new structure was uncovered. If the falsifier is only used to protect the theoretical core of the postulates and the Black Hole status of its singular corroborative events, this is *a degenerative research program*. The recent discussions on how to interpret direct replications are revealing about the willingness to be progressive or degenerative. (compare Brandt et al., 2014; Cesario, 2014; Dijksterhuis, 2013; Simons, 2014; Stroebe and Strack, 2014)

Figure **1.7** displays what the results might be if a more formal system of theory evaluation was adopted in the soft sciences, one that allows positing theories of construction that can be evaluated in progressive research programs and eventually can make point-value predictions about parameters as a theory of principles. In terms of corroboration, a track record of predictive power and empirical accuracy has to be established by a theory. The goal of scientific explanation should be unification, finding out how entities are the same, or can be understood by more fundamental, or sometimes more general laws, a progressive program. Einstein understood however that a unified theory'*pays for its higher logical unity by having elementary concepts* [...], *which are no longer directly connected with complexes of sense experiences.*' (cf., Seevinck, 2011). To create a theory of principles means sacrificing a kind of explanation that is more a personal understanding of the world in terms of everyday experiences. This was not just a coincidental oddity for Einstein, he firmly believed that the fundamental notions of physics cannot be induced from experience and they cannot be justified a priori on the basis of our faculty of knowledge. This strongly opposes Kant's notion of incommensurability between Biology and Physics (e.g., Hasselman, Turvey, Seevinck & Cox, 2011;

Seevinck, 2011).

There is a peculiarity about the role of ontology in this loss of explanatory power that I will leave for thorough discussion in the final chapter of this book in which I will also suggest how to (meta-theoretically) deal with its loss. It has to do with overcoming the enormous psychological satisfaction that is felt when empirical data agree with a cultural belief system or a personal world-view (Salmon, 1990). The danger is that an evaluation of the theory that predicted the data in a more or less formal chain of deductions may be positively biased towards such beliefs by its satisfactory agreement. This danger is quite literally quoted in the definition of the hard/soft science divide by Fanelli (2010): Socio-cultural and psychological factors decide how data are collected, analysed and interpreted. I call this *the interpretation fallacy of theory evaluation* and it occurs very often in the social sciences, but is hardly ever noticed as problematic (Lebel & Peters, 2011 use ‘interpretation bias’ in a slightly different, but related meaning).

1.5.3 Blinded by downward seepage? Cause and effect of the interpretation fallacy.

In the times before modern empirical science the existence-proof of ontological entities to our senses prompted ancient scholars to theorise about their role in the workings of the physical universe and the human body, like the classical elements and the Hippocratic humours. This is a very sensible and scientific thing to do if you have no highly corroborated nomological net to guide you. Theories were constructed based upon their empirical evidence as‘measured’ by our senses. This is essentially related to the logicians explaining implicit empirical content in theoretical propositions, ‘Ramsified upward seepage’ (Meehl, 1990; Meehl, 1978, 2002, 2004b).

Here, we stumble upon the ancient problem of philosophy (mentioned in the preface): Is our internal world, the mind or our consciousness, part of reality? Should phenomena of the mind be allowed as a construct to theorise about the mind or should their existence to our mind’s eye be ignored in theories? We must acknowledge that in social science, but psychological science in particular there is a thin ontic line between the theoretical concept of a mind and the actual sensory phenomenon of mind. Is the line too thin? For example, to the minds eye the external continuous universe appears filled with discrete objects, our internal thoughts, memories of events and the words we use to communicate are all discrete. Therefore, we may be naturally inclined to construct theories based on discrete components, without ever considering continuous architectures of mind as an alternative (Spivey, Anderson & Dale, 2008; Spivey & Dale, 2006).

We experience time as a linear flow, memory, history, prospection, but what if it is not like that at all as some physical theories suggest (Brown & Uffink, 2001; Bunge, 1968; Nottale, 2010). Within the framework of ecological psychology, intentional dynamics and event perception are a very different conception of time and its passage is posited, a perspective known to very few soft scientists (Gibson, 1966; Kugler, Shaw, Vincente, & Kinsella-Shaw, 1990; Shaw & Kinsella-Shaw, 1988; Shaw & Turvey, 1999; Shaw, Flascher, & Mace, 1996; Turvey & Carello, 2012). Perhaps not surprising that the‘inventors’, like roboticists and industrial designers have a different opinion. In the final activity report of the research project MACS (Multi-sensory Autonomous Cognitive Systems Interacting with Dynamic Environments for Perceiving and Using Affordances; Rome, 2008) it is mentioned that a valid starting point for the project was to draw inspiration from cognitive science, being ecological psychology and embodied, embedded cognition, and situated cognition (Rome, 2008, p. 5). Not a description many cognitive scientists would provide of their own field.

As alluded to in the discussion of the historical critiques in the preface, the origin of theoretical diversity and issues with reproducibility in psychology in particular and in the social sciences in general, may be due to fundamental category mistakes. The mistakes concerns identifying theoretical constructs that emerged from human theorising about reality as direct sensory experiences, quite literally observations, due to the thin ontic line that separates them. This is a downward seepage in which empirical phenomena acquire theoretical content. Such theoretical constructs, about which only a scientific theory should be a realist, are mistaken for an actual, or at least a perceptual constituent of reality. As a result the implicit theoretical object or ontology is essentially excluded from scientific inquiry, its existence proof is already provided for by the senses. Something one might call *the empiricist's blind spot*, a blindness induced by downward seepage.

For instance, when an experiment is conducted in which several predictors or independent variables in a linear additive model should explain variance in a dependent variable of interest, there are usually two conclusions drawn when the explained variance is not satisfactory. The first is to add better predictors in the next study. The second is some notion of measurement error that should be resolved by using larger samples or better measurement instruments (i.e., deflecting the modus tollens refutation torpedo towards veracity and validity of theoretical and instrumental auxiliaries). Very rarely one encounters a conclusion that is scientifically equally sound: Wrong model! The variance in this variable cannot be explained by a linear additive combination of predictors or by efficient causation through experimental manipulation. In other words: The wrong causal ontology is used to scientifically describe this phenomenon.

1.5.4 We want components!

An example of such an object that escapes genuine theoretical and empirical inquiry in psychological science, one of the supermassive Black Holes in the nomological net, is the concept of the mental representation (Haselager, de Groot, & van Rappard, 2003; Spencer & Schöner, 2003). It is associated with discrete cognitive architectures and is frequently used to posit hypotheses as an object of measurement or as a vehicle to interpret experimental results. Representations are also attributed specific properties (carriers and/or encoders of information) and are even given powers of causation (Cox & Hasselman, 2013; Hommel, Müsseler, Aschersleben, & Prinz, 2001). Few authors who use the concept however, refer to the theoretical or ontological truth status of the representation, for example by empirically questioning its existence. After careful analysis Haselager et al. (2003) concluded that cognitive science lacks a proper operationalisation of the concept of representation and is therefore unable to discuss whether a system has representations or not. Still, its use continues, far beyond the domain or cognitive science and theories are constructed that have mental representation at their theoretical core, while viable alternative theoretical frameworks including those mentioned above are ignored (cf., Dreyfus, 2002)

To demonstrate how real this phenomenon is and how it causes the interpretation fallacy, an example in which authors explicitly complain about the fact that they cannot interpret predictions and measurement outcomes of theories in terms of their own preferred constituents of reality. A recent commentary by Wagenmakers, van der Maas, and Farrell (Wagenmakers, van der Maas, & Farrell, 2012) entitled: "*Abstract Concepts Require Concrete Models: Why Cognitive Scientists Have Not Yet Embraced Nonlinearly Coupled, Dynamical, Self-Organized Critical, Synergistic, Scale-Free, Exquisitely Context-Sensitive, Interaction-Dominant, Multifractal, Interdependent Brain-Body-Niche Systems.*" The authors evaluate the promise of 15 years of the Complex Systems (CS) approach to cognitive science and claim it has failed to live up to its promise because:

1. phenomena associated with complex systems (such as fractal scaling and self-organised criticality) are "*mysterious phenomena*";
2. Cognitive scientists are not interested in the methods of CS, because they are too "*vague*", "*too general*" and "*mostly about 1/f noise*"
3. Cognitive scientists are not interested in CS because they want to "*infer latent cognitive processes*".

The authors conclude the claims of the CS approach will appear to the cognitive scientist as: "*mostly speculation, wrapped in jargon, inside wishful thinking*". This tendency to explicitly prefer results that can be interpreted as a mechanism of component processes over these '*mysterious phenomena*' is encountered in several articles by this group of authors (e.g. Torre & Wagenmakers, 2009; Wagenmakers, Farrell, & Ratcliff, 2004).

The negative evaluation is based largely on the low explanatory power of these abstract constructs. They do not correspond to the familiar phenomena of human nature in the world-view of the researchers and this should be *irrelevant* in the evaluation of scientific credibility, that is, it the existence of the abstract concepts follows from material implication by observation (corroboration). The interpretation fallacy is most prominently evidenced by the repeated requests to provide mechanistic models and to provide theoretical concepts that can be used to infer latent component processes from empirical findings. At the same time the criticasters do not seem to recognise anomalies in the empirical record to the mechanistic component ontology. Their criticism really seems to be exclusively based on the usage of mysterious theoretical phenomena and very general and vague methods, no references are made to evaluations of the predictive power and empirical accuracy of studies using the CS approach. Such information has been available for a while (cf., Hasselman, 2013) and the number of studies that bring observables like fractal scaling under experimental control in human performance is steadily increasing and should be evaluated for the precision and accuracy of their theoretical predictions based on those '*mysterious phenomena*'. **None** of the following studies, published in or before 2012, were mentioned in the critical evaluation of 15 years of the complex systems approach to cognitive and behavioural science: Correll, 2008[17], 2011; Holden, Choi, Amazeen, & Van Orden, 2011; Holden, Van Orden, & Turvey, 2009; Kello et al., 2010; Kello, Beltz, Holden, & Van Orden, 2007; Kuznetsov & Wallot, 2011; Stephen, Anastas, & Dixon, 2012; Stephen, Dixon, & Isenhower, 2009; Wijnants, Bosman, Hasselman, Cox, & Van Orden, 2009; Wijnants, Cox, Hasselman, Bosman, & Van Orden, 2012; Wijnants, Hasselman, et al., 2012).

In other words, because the researchers evaluation the CS approach cannot interpret the measurement outcomes given by observables posited to exist by theories that describe the dynamics of complex systems, the theory is discarded as vague, uninteresting and void of meaning! In fact, the deep wish of the cognitive scientists to be able to interpret measurement outcomes as a mechanism of components means they want to interpret it as something that corresponds to their daily experiences, a machine, a computer, a manifold of immediate sense experiences. Cutting loose theoretical entities from those experiences means introducing deeper levels of abstraction and this seems a difficult task. Abstraction (and generality) are labelled as problematic quite often when the CS approach is discussed: '*The counterpart of this level of generality is that it is to some extent accompanied by a detachment from the singularity of the phenomenon of interest.*' (Torre & Wagenmakers, 2009, p. 303). I would argue this is not a counterpart, but a benefit and in fact a goal of scientific explanation. The other option is to have a separate theory for each 'phenomenal singularity' and this is exactly what I believe causes the recurring crises in the empirical social sciences.

This is the important existential question the empirical social sciences need to answer: Do they want principled theories that are detached from the singularity of the sensory phenomenon, or not? If they do, Figure **1.7** sketches what may await them: Explanatory power will initially be lost, mysterious constructs will emerge that bear no resemblance to anything perceivable by the senses. If they do not, the crisis will perpetuate. Perhaps the complex systems approach and the responses evoked by its conjectures are a sign of endeavours that attempt to cut the phenomenological ontic umbilical chord connecting the experienced world to theoretical entities. Ecological psychology seems to have attained such levels of abstraction decades ago and are ready to redefine adaptive behaviour from first principles (cf., Turvey & Carello, 2012) .

[17] Madurski and LeBel (2014) recently failed to replicate Correll's study in two samples, with combined N of 296 participants. The replicating authors note however that the time series in this large sample should be characterised as 1/f noise (non-zero spectral slopes). So it was an effect on the scaling exponent that was not replicated.

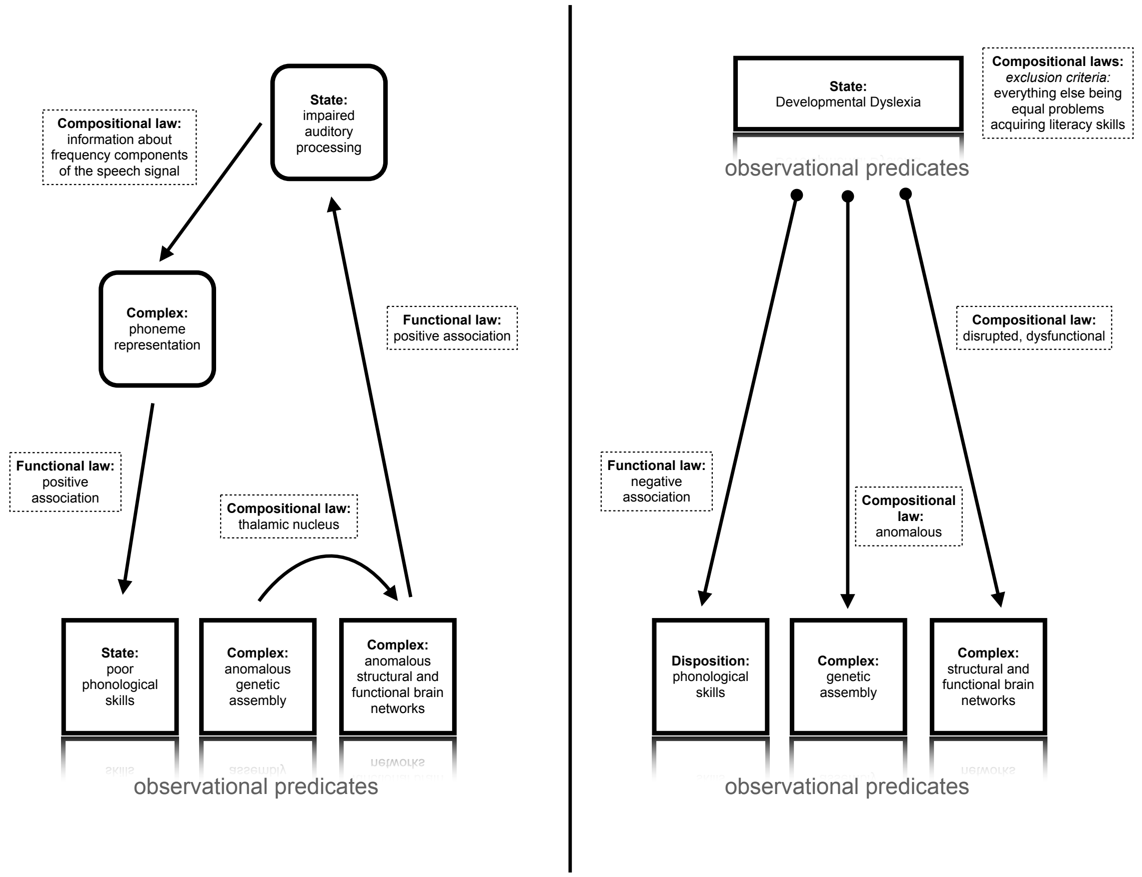


Figure 1.8 – Two possible ways to reconstruct the postulates of an ætiology about impaired auditory perception as a cause of developmental dyslexia.

1.6 An ontology of Failing Components, or a Failing Ontology of Components?

The proclaimed goal of this chapter was to find a theoretical account for the ætiology of developmental dyslexia that could explain the diversity of apparently corroborated accounts evidenced by the analyses abstracts of scientific articles in section 1.2 (the first conjecture). Taking into account the considerations about weak predictions, crud factor and unjust application of the Lakatosian defence a genuine causes of the diversity (the second conjecture), I believe it is sensible to suggest, based on the historical and meta-theoretical analyses I presented that the field is blinded by downward seepage from a (causal) ontology that constitutes a naïve Newtonian physics. Many of the factors discussed contribute to uphold the ontology, especially the 'effect = structure fallacy'. What can we do in terms of meta-theoretical evaluation? First, I will give a practical, and in my opinion honest perspective on the nomological net as it was evidenced and characterised in section 1.2. Then I will explain in what way the empirical inquiries described in Chapters 2-5 of this book can be seen as consequences of adopting this perspective and taking into account most of the issues discussed here. It should be noted however that this is a convergence of evidence rather than a prediction of novel facts.

Figure **1.1** is the best representation of the current causal ontology (analogous figures –with pathways drawn in– and explication of the components can be found in: Ramus et al., 2006; Ramus, 2003a, 2003b, 2004). Imagine the components in the figure are the nodes in a nomological net without laws connecting the entities. Two thirds of the net can be considered the observational subset, the biological and the behavioural components (perhaps with the exception of genetic and environmental factors). The cognitive layer can only be corroborated by inclusion in a derivational chain predicting an observation. The entities represented in the cognitive layer mostly concern states (impaired) and structures that are complexes (representations). When causal pathways are drawn they generally originate from the biological anomalies (structures), connect to one or more cognitive component and to the behavioural phenomena.

What seems to be implied, is a structure like the one in Figure **1.7**, where $\mathbf{A_3}$ is a biological component, $\mathbf{A_1}$ and $\mathbf{A_2}$ are cognitive components and **S** represent behavioural phenomena. This cannot be the case, as the biological anomalies should be closer to **E** because they are observable as direct sensory experiences, or at least closely connected to observational statements. The cognitive components are the most abstract entities and they should have to be 'on top'. What this structure reveals is a world-view in which there is biological determinism, our genes and our brain determine observed behaviour, overt, as well as our mental behaviour (cognition). As a nomological net, the specific graphical hierarchical arrangement with respect to **E** is not essential, nodes can just be said to belong to the observational set or not. It does convey information, as can be seen in the left panel of Figure **1.8**. Some components of the observational set have been placed at the bottom and some laws have been drawn to indicate how an anomalous genetic component dictates the disrupted composition of the thalamic nucleus in the CNS that is associated with the state of impaired auditory processing. This state has consequences for the composition of phoneme representations because the frequency information of the speech signal is depleted. A phoneme representation composed in such a way is associated with the state of *poor phonological skills* observable in a small percentage of children.

1.6.1 Really, we don't need them – Necessity and Sufficiency in interaction.

What is rather odd behaviour by the theorist who deduced the chain or the meta-theorist who is reconstructing the derivation is that the entities are defined with states or modifiers attached (impaired, anomalous, poor). There are many rules suggested that should guide theory (re-) construction, from parsimony and Occam's razor to aesthetics (e.g., Meehl, 2002), but keeping it clean and simple, with as few invisible constructs to start with seems a sound advice. The right panel is an attempt to see what happens if states and modifiers are removed, preferring only theoretical constructs that can D-N predict observations by laws. This representation is not necessarily 'better', or representative of a fundamentally different theory. Perhaps some explanatory power is gained. The cleanest solution in my opinion is to define a state called developmental dyslexia (DD-state); it is part of the observational set by definition due to exclusion criteria. These exclusion criteria take on the role of the ceteris paribus clause in theory corroboration: Everything else being equal, there are children who fail to acquire proficient levels of reading and spelling performance. Nothing more, nothing less.

Whether or not there is an anomalous genetic assembly or structural or functional network of the CNS, conditional on the DD-state, is something that must be corroborated. The same holds for possible partial derivatives, subtypes within meet the requirements by the compositional law. This would only turn the DD-state into a complex of states, or a composition of states, like an order parameter of a dynamical system (which is why I wanted to add that entity to the mentology). In fact, recently Meehl's taxometric methods were used to evidence two separate taxa of developmental dyslexia (O'Brien, Wolf, & Lovett, 2012). It is important to realise that it is this state, by means of its definition, that may lead a theorist to predict that an anomalous composition of structural complexes is associated with it. A state in mentology is defined as *an event spun out over time*. An event is an occurrence where structures interact and change. By declaring something a state, a continuous interaction that results in resistance to change or stability over time, is implied. Therefore the state can'predict' compositions of structures that are believed to be involved in the enduring interaction. As suggested earlier, intervention programs are events that attempt to destabilize the DD-state by tweaking whatever structures a theory posits to be responsible for its continuation.'Merely' positing the DD-state implies deductions are possible about its causes and composition.

What happened to the cognitive components? In short, I really think we don't need them. If something akin to the right panel in Figure **1.8** would be accepted as an honest meta-theoretical account of the current situation. There exist mainly theories of construction that predict signs of associations or anomalous composition of structure. The latter is of the same order of the former because it implies comparing two groups, hypothesising the structural difference to the reference group is non-zero, in the direction of the anomalous group. If I wanted to claim there is a deficit in auditory processing associated to the DD-state that should also be related to poor phonological skills, I predict an association between the two constructs based on the DD-state alone. Phonological representations do not have to be included in the chain to strengthen its logical structure or to hope to improve empirical accuracy. It is hard though, to find theoretical accounts that do not include any notion of phonological representation. One reason may be that they are too close to the thin ontic line and downward seepage from theories of universal grammar and information encapsulation have made them part of the furniture of the world. Chapter 5 will explore this thesis in more detail.

Another reason to keep representations and cognitive modules may be that they are needed as a vehicle for explanation, in order to understand DD-state's negative and anomalous associations to all those other entities. If so, the phoneme representation would be the hub in the network, the entity receiving a lot of corroboration. It would be the link in the derivation chain without which prediction would be rendered trivial. Although it may be used very often, corroboration of its general nature, properties, function, or its vital role in explanation of the DD-state has not occurred to the best of my knowledge (Blomert, Mitterer, & Paffen, 2004; Clark, 1999; Dreyfus, 2002; Farrar & Van Orden, 2001; Haselager et al., 2003; Port, 2010; Ramus & Szenkovits, 2008). I will not proclaim that science does not need a cognitive level at all, but in most cases it feels like dressing up perfectly normal (e.g., crud factor) associations between observables using a level of'scientific' abstraction that serves no formal purpose. Colleagues often object to this statement by saying something like: 'Ok, then explain this twitter discussion we're having without making use of representations' I usually ask in such cases: 'I'll try, but what would be the evidence that would convince you that mental representations do *not* exist?' No answers so far.

1.6.2 Removing the relata: Interaction dominant dynamics.

What happened to the causal pathways? In short, I really think we don't need them either. Indeed, in the left panel one could start at the genetic anomaly and play a game of tag to end up in poor phonological skills. On the right, phonological skills have been redefined as a disposition, something that is realisable (or unrealisable) and will often be attached to a structure by a law defining it as a property or characteristic. When is the disposition realised? This is in fact a complex condition to work out if one adopts the right panel version. I declare a DD-state is defined as a condition of:

1. exclusion criteria of the kind 'all else being equal only this particular should be observable'
2. exposure to falsification, corroboration, amendment and unification, because it is defined as a theoretical entity in a nomological net.

Therefore the set of conditions to evidence the DD-state is a complex conditional. Let Θ denote this set of conditions, when Θ is true, it is can be said to be the cause of the DD-state. However, it is not necessary to infer the state if Θ is met. A 'cause', like the anomalous gene assembly, can be Insufficient by itself, but constitutes a Necessary part of Θ that is itself Unnecessary, but Sufficient to evidence the DD-state (see Mackie, 1965). The condition Θ is called *INUS*.

One example to clarify the *INUS* condition: Suppose a child is tested on the DD-state and an IQ measurement results in 110. Further tests are conducted and the diagnosis is: DD-state. Was the IQ

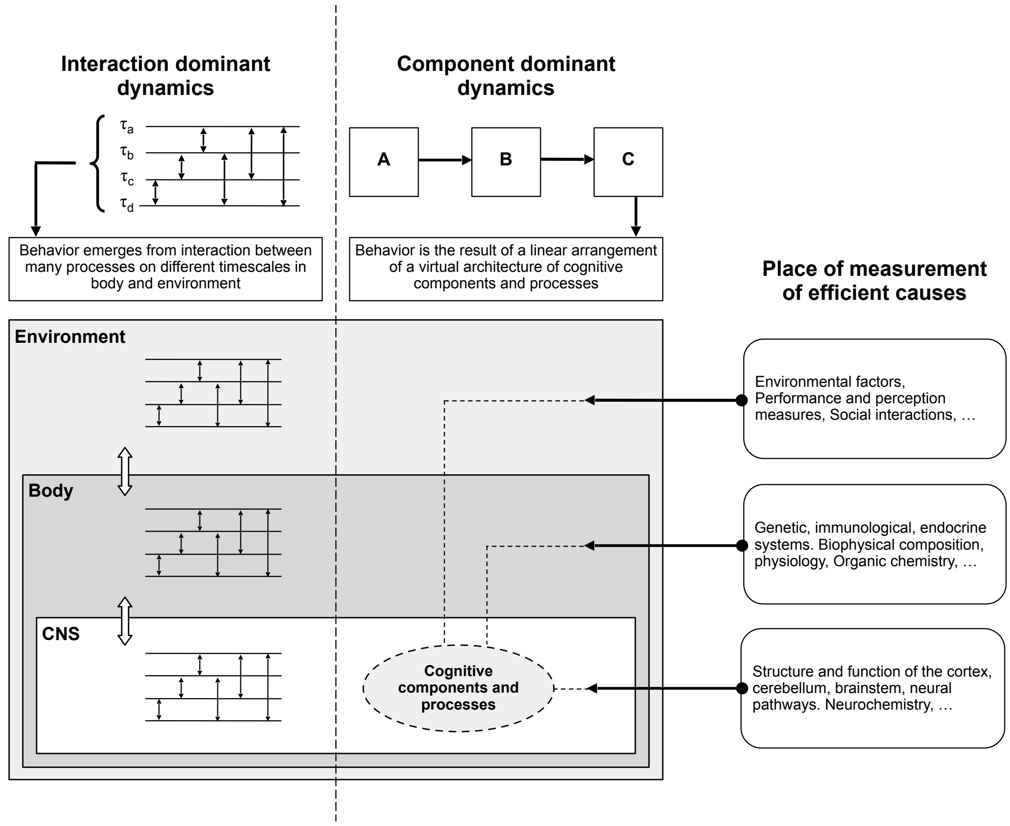


Figure 1.9 – The different ways in which the two ontological frameworks approach the emergence of (impaired) behaviour.

score the cause of the diagnosis? IQ = 110 by itself is not a necessary condition for developmental dyslexia, other causes exist that would lead to the conclusion if IQ were not 110. It is also *Insufficient* as a single condition, not everyone with IQ = 110 is a dyslexic reader. IQ is a *Necessary* part of the complex of conditions that is *Sufficient* to evidence a DD-state. This specific complex condition with IQ = 110 as one of its values is *Unnecessary* to evidence developmental dyslexia. Conditions with other values for IQ or without actual quantitative measurements of IQ, can'cause' the same diagnosis as well. Anyone who wishes to evidence causal pathways should at least consider the fact that the conditions for evidencing the DD-state are a complex of interacting conditionals. Any clinician who has had to diagnose a client will know exactly what I mean by'complex interaction of conditionals'.

By declaring the compositional law for the DD-state as a ceteris paribus clause, one would expect that any corroborating evidence of anomalous behaviour or structure would expand the list of necessary parts of sufficient conditions to evidence the state.'All things being equal' would stop to have any meaning, because everything about the DD-state would be different from everything else. Although the scientific record appears to indicate almost everything is different about dyslexic readers, these things are not added to the INUS condition, save perhaps the recent additional requirements (in the Netherlands) of fluency and response time latencies (Blomert & Vaessen, 2009). Maybe the reality of clinical practice is a better corroborator of verisimilitude than science in this case. Most conditionals suggested in the literature that goes beyond evidencing language and literacy related performance measures are Insufficient by themselves, and Unnecessary for being considered a part of a complex conditional for developmental dyslexia.

The causal structure that would be represented by a nomological net consisting mainly of structures that are complexes –like in the social and life sciences– and on complex *INUS* conditionals –like in the social and life sciences– wouldn't have anything to do with the implied component dominant dynamics of contemporary causal pathways (cf. Van Orden, Holden, & Turvey, 2003). Familiar'degrees of causation', or entailment are possible in component dominant dynamics, such as uniquely explained variance, beta weights or effect sizes. In general, a linear arrangement of partial causes always neatly sum up to produce the behaviour of interest. An alternative causal ontology is interaction dominant dynamics in which not the components themselves, but their interactions as a whole are the source of the observed behaviour (Ihlen & Vereijken, 2010; Kello, Beltz, Holden, & Van Orden, 2007; Van Orden, Holden & Turvey, 2003; Van Orden, Holden & Turvey, 2005; Wijnants, Cox, et al., 2012). Here the contribution of components is not additive, but multiplicative and nonlinear (Holden et al., 2009; van Rooij, Nash, Rajaraman, & Holden, 2013). Such interaction dominant dynamics render individual component behaviour (which are still posited to exist), such as poor performance on ability X, impaired representation of that feature Y, as a less interesting object of theoretical and empirical inquiry.

As a consequence, theoretical and empirical inquiry is aimed at identifying and understanding the contexts in which impaired behaviour emerged. Adopting such a perspective entails that all observable behaviour can only be understood relative to the context in which it was observed, that is, the measurement context (cf. Holden, Choi, Amazeen, & Van Orden, 2010; Van Orden, Kello, & Holden, 2010). Figure **1.9** presents the fundamental differences between the two ontologies in their assumptions about the causes of behaviour and their assumed place of measurement. Figure **1.9** may reveal why the nature of cognitive components and processes remain elusive in their causal role. They are inferred, not postulated, based on data from different places of measurement. Their causal structure does not incorporate the nested nature of both measurements as well as posited entities. Applying the concept of the complex conditional reveals hierarchical dependencies of one condition on another and such a complex, if it were composed of the correct conditionals, should be considered as a whole. As a consequence, impaired behaviour should be understood as emerging from the whole of constituent components, not from an individual component. The notion of a cause is somewhat more radical than the complex conditionals and is known as impredicative, circular causation (Chemero & Turvey, 2010; Freeman, 1999; Turvey, 2007), or nested causation.

1.6.3 Beyond the boundary: Investigating degrees of theory specification

The Chapters 2-5 in this book concern empirical and simulation studies on the role of speech perception and its hypothesised causal role in ætiologies of developmental dyslexia. The studies take the perspective of interaction dominant dynamics to study the origin of the emergence of impaired behaviour. Moreover, I will attempt to take into account many of the critical comments made in this chapter about risky predictions, testing competing theories, the strength of derivation chains and proposing principled theories and'deeper lying entities'. In a sense the endeavour will concern steps of increasing theory specification by increasing the credibility hurdle each time, by producing predictions about the interaction-dominant structure that are more risky than previous predictions. In doing so, the 'similitude' of competing theories can be identified. This is, in my opinion, the only way to deal with theoretical diversity in a science of human nature.

Box **1.4** list a number of steps in theory specification that allow assessment of the similitude of theories in explain the same phenomena, ultimately leading to a decision on their verisimilitude. I believe it is warranted to claim that a proper decision on verisimilitude of theories, either under conditions of strong inference or during evaluation of precision and accuracy of predictions, cannot be achieved in Stage A and Stage B.

Stage A has been the subject of the last paragraphs in which I declared the theoretical entities of the framework that I will use to study developmental dyslexia, the perspective of interaction dominant dynamics, or the complex systems approach to impaired behaviour. What also became clear in previous paragraphs is that simply declaring the formalism, nomological framework, or theory, even if it is based on convincing and logically consistent derivation

Box 1.4: Three different stages can be identified in degree of theory specification. The steps that allow assessment of the similitude of theories are adapted from Meehl (1990a; 1990b; 1990c; 1997).

Three Degrees of Theory Separation

A Scientific claim, either by construction, or by declaring conjectures to principles:
- I Type of entity postulated (substance, structure, event, state, disposition, field)
- II Compositional, developmental, or efficient-causal connections between the entities in **I**

B Tests of associations between entities (**VI** is the common FactorXCovariate interaction)
- III Signs of first derivatives of functional dynamic laws of connections **II**
- IV Signs of second derivatives of functional dynamic laws of connections **II**
- V Signs of mixed second order partial derivatives (Fisher "interactions") of connections in **II**
- VI Ordering relationships among the derivatives in **III**, **IV**, **V**

C Prediction of functional form and parameter values, tests of universality:
- VII Function forms (e. g., linear? logarithmic? exponential?) of connections in **II**
- VIII Trans-situationality (Context relativity) of parameters in **VII**
- IX Quantitative relations among parameters in **VII**
- X Numerical values of parameters in **VII**

chains of truth-like postulates found in the literature, other scientists can to some degree choose to raise a brow and proceed with business as usual.

Stage B reflects the swamp of unstructured empirical facts and therefore theoretical diversity in which the majority of social science is currently 'stuck'.

Theoretical diversity may now be redefined as the indeterminacy of the degree of theoretical similitude. Why is the field stuck? Predominantly, linear models are used to test signs of associations between first and second order derivatives of observables (dependent variables). Step VII of **Stage C** in which a functional form should be investigated is fixed at 'linear'. This does not provide enough specificity of competing theories in order to decide how similar they are.

'Derivative' in the context of an observational constraint as introduced in paragraph 1.3, really just means that different levels of the variables that make up the constraint have been measured and that the differences between those values with reveal something about their systematic covariation: Taking a derivative = taking the difference between values that are ordered according to some principle. If the ordering principle is time, that is, measurement occasions reflect temporal order, a time derivative is obtained. Time takes on the role as a variable to which change can be compared, so time can be one of the variables in the observational constraint. The best-known example is the time derivative of displacement, being velocity, and change in velocity is the second order time derivative of displacement, acceleration. If time is not the variable to which variation is compared, physically speaking, it must be space. This can be interpreted very generally and a sample of participants measured at 'one' occasion, is a 'sample space'. Note that frequency distributions of real-valued random variables have 'moments' of a certain order. The first moment is the expected value of the random variable, often estimated by the arithmetic mean of a finite sample of different values of the variable. The variation in the mean is the second moment, the variance (or standard deviation); third and fourth moments are skewness and kurtosis-like properties, all indicators of change of higher order moments.

The analogy is somewhat more complicated in reality (in addition, the analogy only holds when the ergodic condition applies (Kievit, Frankenhuis, Waldorp, & Borsboom, 2013; Molenaar & Campbell, 2009; Molenaar, 2008; Petersen, 1996), which ensures the time and space averages of variables measured from physical systems will be equal (given infinite time). For current intents and purposes, any linear statistical model that tries to evaluate whether unique changes in one variable are associated with unique changes in another variable, measured at the same point in time, is taking the derivative of a linear function of the values in the sample space. In addition, evaluation of unique variation when the model contains more than two variables means that co-variation of some variables is evaluated at fixed levels of the other variables. This is what is meant in **VI** by taking a partial derivative: *The derivative of the dependent variable with respect to the unique predictor, with the other predictors held constant. Held constant* usually means with respect to the expected value of the other variables in the model (this will depend on the modelling strategy). The same holds for simple FactorXContinuous interactions in which the co-variation of a continuous predictor with the dependent variable is evaluated for the levels of the factor. The interaction between Age and Gender when predicting Reading Performance is a second order partial derivative of a linear function. It is evaluated as a *difference* between cov(Reading,Age) fixed at 'Boys' and cov(Reading,Age) fixed at 'Girls'. A first order partial derivative of a linear function between Y and Gender is a simple t-test of the variation of Y over the sample space, with values fixed at the levels defined in Gender. A *difference* between group means.

The point is, under assumption of the linear form of the functional relationship between random variables, many partial derivatives, even though they are of second order, come down to predicting the sign the difference: mean(Sample1)-mean(Sample2). The fact that this does not suffice to end theoretical diversity is reflected by the dearth of instances of degree **VI** in empirical social science: It seems often to find out which one out of several corroborated derivatives (let's call them effects), is more important. Many historical examples exist, for instance in deciding on the primacy of different fundamental Gestalts in visual perception, but a more recent example is contextual dependency of the observation of the highly corroborated word frequency effect in reading (Bosman et al., 2013; Van Orden, 1987). In a recent study, 13 well-known effects were replicated in more than 60 independent laboratories around the globe (Klein et al., 2014). Interestingly, the effects that were successfully 'replicated' (that is, the sign of the effect was greater than 0), were the ones with highest variability between different samples in terms of the estimate of their magnitude. The only relatively uniform effect size magnitude that was recovered was 0, the effects that did not replicate. From the perspective of nonlinear functional forms, more specific nonlinear dynamics, such variability would likely be part of the prediction (cf. Farrar & Van Orden, 2001).

The progressive step to take, to get out of Stage B into Stage C, would be to define a functional form (**VII**) for the hypothesised connections between entities and test its predictions. The most problematic aspect about fixing linearity for practically the entire spectrum of phenomena associated with human nature, is that scientist are not aware, or, know why, they assume this functional form by default. That is, they do not know other options are available, or do not wish to explore other options due to perceived difficulty of those options. This practice of limited theory specification of an entire field of science does not qualify as a progressive research program from the perspective of a natural science, or any other measure of comparison. In general, when the assumption of independent measurements is found to be invalid, there will be interaction dynamics at work. There might still be linear dynamics behind the derivatives, but even changing the focus of empirical inquiry to answering that question would be advancement out of the swamp that is stage B.

Stage C is the degree of specification at which for instance the physical sciences currently operate. The example I gave of corroborated measurement outcomes predicted by QED, but also the measurements at the Large Hadron Collider that evidenced the Higgs Boson, they are all numerical predictions of parameters (step **X**) of functional forms (or large assemblies of functional forms, i.e., models) of dynamical or compositional lawful relations between theoretical entities. Another way to distinguish between Stage B and Stage C is that Stage B is about fitting model parameters to data, but stage C is about investigating whether reality fits with model parameters, or, simulating reality as specified by the theory.

1.7 Beyond the boundary: The role of speech perception in developmental dyslexia

The chapters that follow can be categorised with respect to the degree of specification with which theoretical constructs are put to the test. Most of these concern an inquiry into the specification of linearity (**VII**), a test of the validity of the assumption of the linear functional form. Inquiries will be about the relationships between theoretical constructs posited by theories and the observed impaired reading performance in developmental dyslexia (e.g., phoneme representations, constituents of the speech signal, theoretical processes and components assumed to play a role in reading and emerging literacy). Whether or not behaviour is observed as impaired depends, statistically, on the negative sign prediction of the difference between samples AVERAGE – IMPAIRED. I will use $\delta < 0$ to denote an 'impairment effect'. This test may be biased if a crud factor, or ambient correlation exists, in combination with selection in the lower tail of a population distribution (Figure **1.5**). There are several ways to claim, theoretically, a predicted observation of association corroborates a posited causal structure rather than the crud factor:

Prospective prediction of the impairment $\delta < 0$ from a state in which the impaired variable is not an observable of the system. This is a specification of mixed derivative of time, reading, and the variable used to make the prediction (**VII**).

Control over Context Relativity: Coherent explanation or control over context relativity of effect corroboration. That is, the appearance and disappearance of an association among variables, or, effect (VIII). The necessity for taking this step in specification may be due to the inability to establish a rank order of effects that are associated with $\delta < 0$ (**VI**).

Strong Inference achieves the same as gaining control over context relativity of the inference of an effect. The inability to decide between different theories predicting effects in one or more different contexts, but not all, can be solved by fixing the context such that the variation in what the theories predict is maximised. If it is not possible to create a context that achieves at least some divergence of prediction, then the theories are similar for all intents and purposes.

Principled Prediction Simulation of principles in which the (system of) functions, parameters and formalised entities are cast into a computational model or other formal system (calculus) that can produce numerical predictions that should be related to observables. This is at present the closest empirical social science can get to produce Einstein's principled theories. The goal is to test whether the parameters of the model can be assumed to be a part of reality by evaluating the empirical precision and accuracy of simulated reality.

Thus categorised, the chapters constitute a specification of the observables derived to play a role in the co-variation of speech perception ability, reading ability and the state 'developmental dyslexia', from theoretical entities that are based on interaction-dominant dynamics as an ontology for behaviour and the methods and metaphors that describe the dynamics of complex systems and networks.

Chapter 2 - Context relativity of ordering relationships and prospective prediction:
Predicting Reading Performance from Pre-Literate Speech Perception in Children at-risk for Dyslexia

Chapter 2 reports of a longitudinal study of children at risk for dyslexia and examines how their speech perception performance under different levels of perturbation (according to the auditory temporal processing deficit hypothesis) is related to their reading ability one year later (prospective prediction). Results are explained in terms of measurement contextuality (Barrett & Kent, 2004; Hermens, 2011) and (multi-)stability of observed performance (Farrar & Van Orden, 2001).

Chapter 3 - Principled simulation of context relativity:
When opposites attract, repel and deceive: Using Recurrent Neural Computation to Model Multistable States.

The results obtained in Chapter 2 are modelled using a recurrent neural network based on the principle of behaviour emerging from the interactions of processes on different spatial and temporal scales. This is the metaphor of the state space of a complex dynamical system and it is suggested that the empirical results of participants of which $\delta < 0$ was established in Chapter 2 can be explained as attractor dynamics in a destabilized state space.

Chapter 4 - Strong Inference:
Classifying Complex Dynamic Patterns Into Phoneme Categories

One of the critiques emerging from the meta-theoretical and historical analysis so far is that theories do not confront each other under conditions of strong inference (Hasselman, 2013; Platt, 1964). In addition, deeper-lying entities are not posited, and not tested. Often each theoretical account confirms its own predictions as if a direct comparison were conducted. Chapter 4 examines claims of two'temporal' auditory processing theories, one claiming information encoded in spectral features of the speech signal change too fast to be perceived, one stating the amplitude envelope changes too slow to be perceived by dyslexic readers. The third claim tested is based on an interaction-dominant account of speech perception that states the features of the speech signal listeners use to categorise sounds are collective variables, like those described in synergetics (Akhromeeva & Malinetskii, 2009; Haken, Kelso, & Bunz, 1985; Turvey, 2007). Here these collective variables are extracted from the speech signal as the dynamical invariants of a reconstructed phase space that is assumed to represent the interaction between perception and action cues in the speech signal.

To keep the context fixed and vary the predictions, the variables deemed important by the other theoretical accounts were extracted from the same set of stimuli; there were no different datasets, for different theoretical predictions. The ability of a simple classifier to recover the classifications of speech stimuli by average and dyslexic readers, based on the different theoretical features that were extracted was evaluated as corroboration of a prediction by a theory.

Chapter 5 - Principled simulation of posited entities and strong inference:
Beyond the Static Phoneme Boundary: The Nonlinear Dynamics of Emerging Literacy

The results from previous chapters are interpreted in terms of an interaction dominant coupling hypothesis of the emergence of $\delta < 0$ over time. A general model is proposed that represents the potential landscape of two interacting, coupled collective variables. The coupling strength defines the interaction dynamics and is proposed to co-vary with age. The model predicts nonlinear dynamics that question the existence of static phoneme boundaries posited as entities by several competing ætiologies. Also, destabilised internal structure of phoneme categories due to reduced coupling strength in dyslexic readers is predicted to have an effect on the observed dynamical patterns. Coupling strength is also hypothesised to increase with age, in average readers as well as dyslexic readers. In addition, the model explains how a recently proposed speech perception ætiology of developmental dyslexia (allophonic perception hypothesis) may have erroneously inferred additional speech sound categories to exist as discrete structures, or perceptual boundaries. The coupled potential model provides an explanation that does not include positing additional entities to explain the same phenomenon.

Chapter 6 - A General Discussion of Principles
The Role of Internal Representations in Ætiologies of Developmental Dyslexia

The final chapter will provide an integrative discussion of the results and suggest directions for future empirical inquiries based on the development of a formalism based on the principles and physical laws of the adaptive behaviour of complex systems. This formalism will allow a conception of computation and re-presentation of behavioural modes as the result of an order generating process. The difference between meaning and information is discussed and it is concluded that the component-dominant and interaction-dominant causal ontologies both describe changes in the amount of information necessary to describe the states of the complex living system (or: increase of entropy). These changes are due to the emergence of order in a complex system and current plausible physical and biological explanations of this phenomenon are provided by the sciences that study the adaptive behaviour of open complex systems that exist far-from-thermodynamic equilibrium.

Notes **Chapter 2: Ordering Relationships & Prospective Prediction**

- This chapter may be cited as:

  Hasselman, Fred (2014): Predicting Reading Performance from Pre-Literate Speech Perception in Children At-Risk for Dyslexia: Context Relativity of Ordering Relationships and Prospective Prediction. In *Beyond the boundary. An analysis of verisimilitude and causal ontology of scientific claims: Ætiologies of developmental dyslexia as a case in point fig***share**. *http://dx.doi.org/10.6084/m9.figshare.928517*

- *The supplemental materials to this chapter may be cited as:*

  Hasselman, F., (2014). Beyond The Boundary - 2nd Chapter: Supplemental Materials. Retrieved from Open Science Framework, https://osf.io/9mtpq

- *Parts of this chapter are taken from the manuscript:*

  Hasselman, F., & Verhoeven, L. (unpublished). Context Relativity of Impaired Performance: Stability, not Accuracy of Speech Perception is Systematically Associated with Reading Fluency in Developmental Dyslexia (Sometimes). Learning and Individual Differences.

Chapter 2

Predicting Reading Performance from Pre-Literate Speech Perception in Children At-Risk for Dyslexia

Context Relativity of Ordering Relationships

An important notion that needs to be introduced in support of the plea for an evaluation of interaction dominant dynamics as ontology of (impaired) behaviour introduced in the previous chapter is the difference between the stability of a performance and its accuracy. Although performance may be accurate, for example, a word is read aloud correctly on a reading test, a distinction can and should be made between stable, proficient, coordinated performance and unstable performance leading to that response (see Farrar & Van Orden, 2001; Van Orden, Kloos, & Wallot, 2009; Wijnants, Bosman, Hasselman, Cox, & Van Orden, 2009). The analogy of a path may be useful here: Some paths are easily traversed and get you straight to where you want be (the correct answer), other paths will get you there as well, but may take more time or more effort. In a similar fashion different paths may lead to erroneous answers as well. Exactly which path will be traversed is dependent on the context in which someone is asked to choose a path in the first place. In the context of psychological measurements, the one being asked to choose a path, that is, the participant is a context as well. A theory about both proficient and impaired performance pertaining to a skill like reading should take into account the fact that a scientific theory only makes definite predictions relative to well-defined measurement contexts (see Box 1.1). Such a theory of measurement is not readily available in psychological science as it is in the physical sciences (see e.g., Cox, Hasselman, & Seevinck, 2011). Therefore, in the present chapter we will focus on a detailed task analysis of variables that are hypothesised to be causally related by the speech perception deficit hypotheses of developmental dyslexia. This analysis will show just how well such contexts need to be defined.

2.1 There is a difference between knowing the path and walking it

When comparing empirical results from systems that may be organised in fundamentally different ways, as may be the case when comparing average readers to dyslexic readers, being aware of measurement contextuality becomes rather important. For one, the system on which measurements are performed is part of that context and finding differences in proficient behaviour between healthy- and patient-groups can be expected a priori and may thus be theoretically less informative than currently believed (cf. Farrar & Van Orden, 2001; Van Orden, Pennington & Stone, 2001; VanOrden & Kloos, 2003). What may be more informative as a test of the predictive power of a theory is to experimentally create theoretically prescribed measurement conditions in which behaviour of healthy participants becomes less proficient and more like patient behaviour. This is rarely done, sometimes because of obvious ethical considerations. Contrariwise, attempting to make the behaviour of patient-groups become more proficient, is of course much more common and often a goal of research programs. In the context of developmental dyslexia this would refer to an intervention study with a pre- and post-treatment measurement of reading and spelling performance. Many intervention programs are based on the idea of impaired, inaccessible or fuzzy phonological representations or codes as the primary cause of impaired reading. Indeed, some try to ameliorate reading performance by making these sound representations better accessible or less fuzzy (e.g., Elbro, Børstrom & Petersen, 1998; Tallal, 2004). One important and still hotly debated consequence of this supposed cause is that dyslexic children must show some form of deficit in speech perception (see Serniclaes, 2006 for an extensive discussion of the speech perception deficit hypothesis). In the study presented in this chapter, I will explore the theoretical claims about the relation between speech perception of pre-literate children with a familial risk for dyslexia and their reading performance one year later. The context I will use is the controversial auditory temporal processing deficit hypothesis that attributes to the supposed impaired speech perception of dyslexic readers an underlying impairment of low-level auditory processes (Farmer & Klein, 1995 is a review of this hypothesis). I explicitly note that it is not my purpose to seek empirical support for, or for that matter against, this account as a viable explanation of developmental dyslexia. What this chapter will hopefully show is that measurement contextuality needs to be taken into consideration when evaluating tests of (cognitive) performance (see Holden, Choi, Amazeen & Van Orden, 2010; Van Orden, Kello & Holden, 2010) and more specifically, when theorising about and directing experimental attention towards the causes of impaired reading performance.

The auditory temporal processing deficit hypothesis suggests speech stimuli with rapid transitional elements are processed less accurately because such elements occur too fast to be perceived by those afflicted with the processing impairment. In this context, it has been claimed that acoustic manipulations of the speech signal may facilitate speech perception. Tallal and Piercy (1973, 1974, 1975), who attempted to facilitate speech perception in children with Speech Language Impairments (SLI), introduced such manipulations. The results of these and other experiments led to the auditory “temporal processing” deficit hypothesis, which was subsequently expanded from a theory of SLI to account for the reading and spelling impairments of developmental dyslexia as well (Farmer & Klein, 1995; Habib, 2000; Stein, 1993; Tallal, et al., 1993). The rationale behind the acoustic manipulations is that if dyslexic readers cannot process the rapid spectral transitions in speech, they may therefore benefit from not only a slowing of the speech signal (Tallal & Piercy, 1975) but also amplification of the fast transients (Nagarajan, et al., 1998) or both manipulations (Tallal, et al., 1996; Tallal & Patterson, 2000). Nevertheless, the facilitative effects of these manipulations are highly debated. According to Ramus (2003), the initial studies were not double-blind placebo controlled. The results of more recent studies, including RCT designs have shown that the training program is not any more effective than traditional intervention programs for both dyslexic readers and speech language impaired children (Cohen, et al., 2005; Friel-Patti, Loeb, & Gillam, 2001; Gillam, Loeb, & Friel-Patti, 2001; Gillam, et al., 2008; Hook, Macaruso, & Jones, 2001; Loeb, Gillam, Hoffman, Brandel, & Marquis, 2009). In several studies using manipulated stimuli, Habib and colleagues nevertheless report positive effects on the auditory, phonological, and literacy skills of dyslexic readers of French (Habib, 2003; Habib, et al., 1999; Rey, De Martino, Espesser, & Habib, 2002). In yet another training study with manipulated stimuli, Agnew, Dorn and Eden (2004) found only effects on the speech discrimination of children with dyslexia and no generalisation to their phonological awareness or pseudo-word reading.

Moreover, there is also evidence, however, that it is not just the rapidity of the stimulus or stimulus elements that causes poor perception. When speech sounds are compared to sine-wave speech analogues, perception deficits in dyslexic readers are only found in response to the speech stimuli, not the sine waves (Mody, Studdert-Kennedy & Brady, 1997). Moreover, when synthetic speech is compared to natural speech, impairment is only found for the perception of the synthetic stimuli (Blomert & Mitterer, 2004).

In the present study, the question of the existence of a speech perception deficit in developmental dyslexia is approached in terms of the ability of the system to generate stable patterns of behaviour. The hypothesis is very general in terms of neurobiology as the assumption is made that anomalies such as perisylvian ectopias, planum temporale asymmetry, mycrogyria and thalamic distortions (e.g. Eckert, 2004; Eckert & Leonard, 2000; Galaburda, Menard, & Rosen, 1994; Stein & Talcott, 1999) may destabilise the activation patterns necessary for proficient, fluent, well-coordinated performance of speech perception, phonological awareness tasks, reading aloud, etc., but not necessarily the accuracy of that performance. The specifics of the measurement context, such as the need for a speeded response, the type of stimulus presented and the difficulty of the task itself, serve as constraints that influence the ability to generate stable performance. They decide which paths will lead to the correct answer, which paths are blocked how many detours need to be taken. Such contexts determine whether impaired performance will emerge as unstable yet accurate, or unstable and inaccurate behaviour.

2.2 How to Study Stability?

Colangelo, Holden, Buchanan and Van Orden (2004), introduced a way to analyse response-time dispersions as a measure of stability of performance. In their study aphasic patients and students performed word-naming and lexical-decision tasks (i.e., decide whether a word presented on the screen is an existing word or a pseudo-word). The dispersion of response times was found to be indicative of the quality of performance on these tasks. The correlation patterns of the standard deviations of the response times on the lexical decision task and the word-naming task with the measures of lexical-decision accuracy and word-naming accuracy, respectively, were found to be virtually the opposite of each other for the student and patient groups. In unimpaired individuals, a stability of performance, which resulted in consistent relations between lexical-decision accuracy and word-naming accuracy, occurred. No relationship between the dispersion of the response times (SD) and accuracy was found. In patients with major language problems, instability of performance did not yield consistent relations between the two accuracy measures, whereas consistent relations occurred between SD and the accuracy measures. This finding strongly suggests that there is more to performance than just its accuracy. In a group that does not show any impairments of behaviour we observe that higher stability and consistency with which the correct answer is reported (expressed as a low dispersion of response times) reveals proficient and well-coordinated behaviour by means of a lack of correlation between the two measures. The opposite is true for the group of participants with impaired behaviour: Individual differences in proficiency of performance are expressed by systematic relations between response-time dispersion and accuracy. This relation between proficient (accuracy) and well-coordinated behaviour (stability, speed) has been shown to exist across many tasks, modalities and (time)scales (see e.g., Wijnants, Bosman, Cox, Hasselman, & Van Orden, 2011).

To explore similar hypotheses in the context of development of impaired reading and speech perception, simple speech-perception experiments were conducted. Normal and manipulated stimuli were presented to Dutch kindergarteners with a familial risk for dyslexia and typically developing peers in an identification and a discrimination paradigm. System instability may then be quantified as the observation of correlations between response-time dispersion and the accuracy of performance, whereas the absence of such a correlation would be indicative of stable, proficient performance. An additional hypothesis concerns the relation between stability of performance in speech perception and reading fluency. Since in most transparent orthographies reading performance (and thus diagnosis of reading delay or impairment) is measured using a timed reading test (i.e., number of words read aloud accurately within one minute), reading performance on such fluency tests is expected to be associated with the stability of speech perception performance, in dyslexic readers, but not in average readers. However, if a measurement context in a speech-perception task allows for dyslexic readers to generate stable performance (i.e., by being a very easy task) it may be expected that associations disappear for that particular (easy) context.

To evaluate the relations between the speech perception skills of children and their reading performance (as reading fluency), speech perception was assessed when the participants were in kindergarten (i.e., pre-literate) and they were measured again after they had received one year of reading instruction. The different groups (familial risk vs. no risk in kindergarten; reading impaired vs. average reader one year later), the stimuli (acoustic manipulations of /bAk/ vs. /dAk/) and speech perception tasks (identification vs. discrimination) were treated as different measurement contexts in which different correlation patterns between response-time dispersion and accuracy of performance are expected. As for the different groups, the familial-risk group is expected to display most unstable speech-perception performance as indicated by a larger dispersion of response times. Due to the rather simple task, a group difference in the accuracy of performance is unlikely to be observed. An obvious difference between the two perception tasks is that speech discrimination requires a speech sound to be kept in memory for a short while for comparison purposes whereas identification does not. Short-term memory load has been identified as a task demand that elicits impaired performance on the part of dyslexic readers (Ramus & Szenkovits, 2008), a speech-discrimination task may elicit more unstable performance on the part of such respondents than a speech-identification task. As for differential effects of the acoustic manipulations on speech-perception performance (i.e., not as an intervention), the empirical support for facilitation appears inconclusive, though some authors report phoneme lengthening to be beneficial (e.g., Rey, et al., 2002; Segers & Verhoeven, 2005; Verhoeven & Segers, 2004). Therefore specific predictions will not be made, but it is expected that the different manipulations will reveal differences in performance stability. Finally, the relation between performance stability of speech perception in kindergarten and reading fluency one year later is expected to be affected by context (group, task, stimulus) as well. The general prediction is that the dyslexic readers as a group (as identified by reading fluency measures) will turn out to, in retrospect, display the most unstable speech-perception performance. Moreover, task and stimulus contexts are predicted to influence the observation of these associations.

2.3 Method

2.3.1 Participants

Participants were 38 children in kindergarten (mean age = 5.3, SD = 0.9); 19 were identified as having a familial risk of developmental dyslexia (mean age = 5.3, SD = 0.9) and 19 had no known history of dyslexia in the family (mean age = 5.5, SD = 0.7). The criterion for being at risk for dyslexia was having at least one parent who was dyslexic. This was tested using the criteria set by Kuijpers et al. (2003) for the identification of adult dyslexia and included a timed word-reading task (Brus & Voeten, 1973), a timed non-word reading task (van den Bos, Lutje Spelberg, Scheepstra, & de Vries, 1994) and the verbal-competence task from the WISC (Wechsler, 1955). All of the participating children had a normal vocabulary measured by a passive-vocabulary test (Verhoeven & Vermeer, 2001). Records showed none of the children to have auditory or visual impairments according to standard screening tests conducted at the school. All of the participants were followed until they had received one year of reading instruction and thus reached the end of first grade. Two reading tests were then administered: A timed-reading task for regular words (Verhoeven, 1995) and a timed pseudo-word reading task (van den Bos, et al., 1994). When the child's scores on both tests were within the 25th percentile, the child was considered to have severe reading problems. This resulted in a group of nine children with reading problems, all of whom came from the group of children with a familial risk of developmental dyslexia.

2.3.2 Speech Stimuli

The stimuli were based upon natural speech recordings for the words /bAk/ [container] and /dAk/ [roof]. The stimuli were created to represent the end points of a ten-step /bAk/ to /dAk/ continuum (Schwippert & Koopmans-van Beinum, 1998; van Beinum, Schwippert, Been, van Leeuwen, & Kuijpers, 2005) using the Praat program (Boersma & Weenink, 2001) and have been used in several studies (Been & Zwarts, 2003; Been & Zwarts, 2004). The two stimuli differ only with respect to the second formant transition. The two stimuli were manipulated in three manners using the Praat program (Boersma & Weenink, 2001). First, amplification of all fast transitional elements by 20dB was undertaken. The algorithm used to do this in Praat was similar to the one used by Nagarajan (1998), who confirmed this in personal communication with Segers and Verhoeven (2002). Second, the speech signal was slowed to 150% of its original length (see e.g., Segers & Verhoeven, 2005). Third, both manipulations were applied as in the FastForWord program (Merzenich, et al., 1996; Tallal, et al., 1996): the speech signal was slowed to 150% of its original length and all of the fast transitional elements were then amplified by 20dB.

2.4 Procedure

2.4.1 Speech perception experiments

Identification and speech-discrimination tasks were presented on a laptop computer in a quiet room at the children's school. The order of presentation for the two tasks was balanced across the two groups. In the identification task, the participants were presented a smiley face on the screen, which then uttered a word. After utterance of the word, two frames with a picture in each appeared on the left and right of the screen. The picture in one of the frames corresponded to the word that was just uttered. The child's task was to identify the correct picture by pressing a designated button on the left or right of the computer keyboard. The pictures presented in the frames were randomly interchanged for each presentation. Prior to the experimental trials, 10 practice trials were presented in which the participants received feedback. During the experimental condition, the three types of manipulated /bAk/ and /dAk/ stimuli and the unmanipulated stimuli were presented in a random order. Each word was presented a total of 8 times resulting in 64 stimulus presentations.

In the discrimination task, two smiley faces appeared on the laptop screen. First one face uttered a word and then, after 500ms, the second face uttered a similar or dissimilar word. The children were told that the second face was trying to say the same word as the first face had said. They were then asked to listen very carefully and decide whether the second word was the same or not. After the utterance of the two words by the smiley faces, two frames with pictures in them appeared on the left and the right of the screen. One of the frames contained a picture of two green smiling faces. The other frame contained a picture with a green smiling face and a red frowning face. The children were told to press a designated key on the keyboard corresponding to the left or right frame: When the two words were the same they had to press the button corresponding to the two green smiling faces and when the two words were not the same the button with the mismatched faces. The pictures in the frames were randomly interchanged upon each presentation. Prior to the experimental trials, 10 practice trials were presented in which the participants received feedback. During the experimental condition, the three types of manipulated /bAk/ and /dAk/ stimuli and the unmanipulated stimuli were presented in same and different pairs (pairs were always of the same stimulus category) with the different pairs presented as both /bAk/ and /dAk/ and /dAk/ and /bAk/.

Each pair type was presented, just as in the identification task, a total of 8 times in a random order, which resulted in 64 presentations.

2.4.2 Computation of Sensitivity and Bias Indices

Hits, misses, false alarms and correct rejections were calculated for each stimulus. In the identification task, it was assumed that the goal was to detect the transient formant transition present in /dAk/. This creates a somewhat artificial classification of responses necessary for the computation of sensitivity and bias indices, but we are assuming here that the problems of dyslexic readers lie with identifying fast formant transitions. Pressing the picture "dak" in response to the sound of /dAk/ would constitute a hit while pressing the picture of "bak" in response to the sound of /dAk/ would constitute a miss. When the sound of /bAk/ is presented, a false alarm would be pressing the picture of "dak" and a "correct rejection" would be pressing the picture of "bak."

In the discrimination task the aim of the task is assumed to be the detection of a difference between the stimuli. This meant that responding "different" to a different-pair constituted a hit, responding "same" to a different pair constituted a miss, responding "different" to a same pair constituted a false alarm and responding "same" to a same pair constituted a correct rejection.False-alarm rates, hit rates and proportion correct responses were calculated for each participant and each condition. Based on these scores, the sensitivity index A' (Aaronson & Watts, 1987; Donaldson, 1992; Grier, 1971) was calculated, which is a nonparametric analogue to the more common d' statistic. A' ranges from 0 to 1, where .5 means chance performance and 1 means perfect discrimination. The formula for calculating A' is given by equation 2.1:

$$A' = \frac{1}{2} + \frac{[Hits - FalseAlarms) * (1 + Hits - FalseAlarms)]}{[4 * Hits * (1 - FalseAlarms)]} \tag{2.1}$$

To evaluate response bias the B''_D (Donaldson, 1992) measure was calculated which is a nonparametric analogue to the c bias statistic (equation refBD). B''_D ranges from -1 to 1, a value of 0 means no bias a value of -1 or 1 an extreme bias for either response option. B''_D is calculated as follows:

$$B''_D = \frac{[1 - Hits) * (1 - FalseAlarms) - Hits - FalseAlarms]}{[1 - Hits) * (1 - FalseAlarms) + Hits - FalseAlarms]} \tag{2.2}$$

I chose to use nonparametric measure of discriminability and bias (in fact, they are logistic analogues) given the relatively low number of trials in each condition. The inclusion of more trials would have increased the risk of confounded results due to fatigue and/or loss of attention on the part of the children in the present study.

2.5 Results

2.5.1 Comparing children at-risk to children not-at-risk for developmental dyslexia

A. Speech-identification task

The proportions correct, A' (discrimination indices), B''_D (bias indices) and response times (see Table 2.1) were entered in a repeated measures 2x4 multivariate analysis of variance with a within-subjects manipulation factor of four levels (No Manipulation, Amplified, Slowed Down, Both) and a between-subjects group factor of two levels (Familial Risk, No Risk). Polynomial linear contrasts were defined to explore trends in the repeated data. For the proportion correct responses, a significant main effect of manipulation was found (*Pillai's trace* = 0.40, $F(3,34) = 7.45$, $p < 0.01$, *partial*

η^2 = 0.40). Neither a main effect of group nor an interaction between manipulation and group was found. To further explore the origins of the main effect of manipulation paired samples t tests were conducted with Bonferroni adjustment of the alpha level. The results showed the manipulation in which the speech stimuli were slowed down and amplified to elicit fewer correct responses than the other manipulations and no manipulation whatsoever (No Manipulation vs. Both: $t(37) = 4.14$, $p < .01$, *Cohen's d* = .98; Amplified vs. Both: $t(37) = 4.09$, $p < .01$, *Cohen's d* = .71; Slowed Down vs. Both: $t(37) = 4.09$, $p < .01$, *Cohen's d* = 1.03).

For the sensitivity index A', a significant main effect of manipulation was again found (*Pillia's trace* = 0.34, $F(3,34) = 5.90$, $p < 0.01$, *partial* η^2 = 0.34). There was again no main effect of group, and the interaction between manipulation and group was nonsignificant. When paired samples t tests with Bonferroni adjustment of the alpha level were undertaken, the same pattern of results was found as for the proportion correct responses: The double manipulation caused the stimulus to become less discriminable for all of the children in the study when compared to the other manipulations or no manipulation whatsoever (No Manipulation vs. Both: $t(37) = 3.22$, $p < .01$, *Cohen's d* = .72; Amplified vs. Both: $t(37) = 3.17$, p < .01, *Cohen's d* = .56; Slowed Down vs. Both: $t(37) = 3.61$, $p <$ .01, *Cohen's d* = .79).

For the response bias index B''_D, a significant main effect of manipulation was found once again (*Pillai's trace* = 0.22, $F(3,34) = 3.24$, $p < 0.05$, *partial* η^2 = 0.22). All other effects were nonsignificant. When paired samples t tests were conducted with Bonferroni adjustment of the alpha level, again only one significant difference emerged: The double manipulation predisposed all of the children towards a response bias in the direction of the detection of a fast transitional element in the stimulus when no such element occurred (e.g., a response of /dAk/ occurs when /bAk/ is presented). This was, however, only the case relative to the manipulation involving a slowing of the stimulus (Slowed Down vs. Both: $t(37) = -3.11$, $p < .01$, *Cohen's d* = .66) . Whether or not the bias index B''_D deviated significantly from zero was next examined for the entire group and for each manipulation. This is of interest as it indicates the direction of any response bias with a B''_D value of zero indicating no bias whatsoever. This results showed only a significant deviation from zero for the effects of double manipulation on the bias index ($t(37) = 3.39$, $p < 0.01$).

The analyses of the response times for the identification task showed only a significant main effect of Group, $F(1,36) = 9.31$, $p < 0.01$, *partial* η^2 = 0.21. In general, the No-Risk group responded faster to the stimuli than the Familial-Risk group. No significant correlations were found between the measures of the children's speech perception in kindergarten and their grade one reading performance.

B. Speech-discrimination task

The proportions correct, discrimination indices A', response bias indices B''_D and response times (see Table 2.2) were entered into a repeated measures 2x4 multivariate analysis of variance with a within-subjects manipulation factor of 4 levels (No Manipulation, Amplified, Slowed Down, Both) and a between-subjects group factor of 2 levels (Familial Risk, No Risk).

For the proportion correct, a significant main effect of manipulation was found (*Pillai's trace* = .33, $F(3,34) = 5.61$, $p < 0.01$, *partial* η^2 = 0.33). Neither a significant effect of group nor a significant interaction between manipulation and group were found. To further explore the origins of the significant main effect of manipulation paired-samples t tests were conducted with Bonferroni adjustment of the alpha level. Two significant differences emerged: The double manipulation produced fewer correct responding than the amplified and slowing manipulations separately (Amplified vs. Both: $t(37) = 2.82$, $p < .01$, *Cohen's d* = .49; Slowed Down vs. Both: $t(37) = 3.71$, $p < .01$, *Cohen's d* = .64).

The sensitivity index A' showed again a main effect of manipulation (*Pillai's trace* = .36, $F(3,34)$

= 6.33, *p* < 0.01, *partial* η^2 = 0.36). There was neither a main effect of group nor a significant interaction between manipulation and group. Paired-samples t tests with Bonferroni adjustment of the alpha level revealed only one significant comparison: Those stimuli which were slowed down were more discriminable than those stimuli which were both slowed and amplified ($t(37) = 3.61$, $p < .01$, *Cohen's d* = .73).

No significant differences were found for the measures of response bias, B''_D. When the response bias was tested for deviation from 0, a significant difference was found for the double manipulation only, ($t(37) = 2.24$, $p < .05$). The response times on the discrimination task showed a main effect of Group ($F(1,36) = 5.79$, $p<.05$, partial η^2 = .14) with the No-Risk group responding faster in general than the Familial-Risk group. No other significant differences were found. No significant correlations were found between the measures of the children's speech perception in kindergarten and their grade one reading performance.

C. Response-time dispersion

Spearman correlations were calculated between the Standard Deviations for the response times to each type of speech stimulus (i.e., three types of speech manipulation and no manipulation) in kindergarten and the children's timed word-reading and pseudoword-reading performance in first grade (i.e., severity of reading problems). This was done for the Familial-Risk and No-Risk Groups, separately. The identification task produced no significant correlations for either group of children. The discrimination task, however, produced high negative correlations for the unmanipulated stimuli and amplified stimuli with both the children's word-reading and pseudoword-reading performance in first grade but only for the Familial-Risk group (see Table 2.3). More specifically, greater response-time stability (i.e., lower SDs) for both the unmanipulated and amplified stimuli in kindergarten was associated with better reading performance in first grade.

2.5.2 Comparing average readers to dyslexic readers

The participants in the present study were next divided into a group showing serious reading problems at the end of first grade (n = 9) versus a group showing no such problems (n = 29) (see Methods). A procedure to analyse the dispersion of the response times for the children with reading problems versus those without such problems as introduced by Colangelo et al. (2004) was applied. A nonparametric variance ratio, which resembles a standard ANOVA, was computed. For each participant and each experimental condition the variance in the response times was computed. These variances were then summed within each group or condition and divided by n-1 to produce a variance statistic similar to the mean square in an ANOVA. For the participants with reading problems, the n-1 was 8; for the group without reading problems, the n-1 was 28. An F ratio was obtained by dividing the two statistics. To decide whether the variance distribution for the group of children with reading problems differed significantly from the variance distribution for the group of children without reading problems, a nonparametric bootstrapping procedure was used (Efron & Tibshirani, 1998).

This procedure had to be followed because the group of children with reading problems proved much smaller than the group without reading problems and the assumption of normality could not be met. The objective of the bootstrap procedure is to determine if the obtained F ratios could possibly be found by chance alone by repeatedly selecting 9 and 29 variances on a random basis from the total of 38 variances and calculating an F ratio as described above after each randomisation. The number of randomly sampled F ratios that equalled or exceeded the F ratios for the non-randomised samples was then divided by the total number of randomised samples. The result is a probability estimate similar to a p-value. The data were re-sampled a total of 10,000 times for each speech perception task and type of manipulation. The obtained p-values for all but the double manipulation fell below .05 (see Table 2.4).

Table 2.4 shows the dispersion of the response times in both the identification and discrimination tasks for those kindergarten children who turned out to have serious reading problems in first grade to be greater than the dispersion of the response times for those kindergarten children who turned out to be average readers in first grade.

2.6 Conclusions and Discussion

The results from the speech perception tasks show differences in the accuracy and discriminability measures of the entire group of children depending upon the type of stimulus presented (i.e., measurement context). There were no significant group differences for these measures. Most striking is the finding that the manipulation used in the FastForWord intervention program, namely the double manipulation involving both the slowing of the acoustic signal and amplification of the fast transitional elements, yielded significantly fewer correct responses than all of the other conditions in both experiments. The sensitivity index A' showed significantly less discriminability of the speech stimulus which underwent both manipulations. The main effect of manipulation found for the response bias index B''_D and the fact that the stimuli which underwent a double manipulation deviated significantly from zero revealed a bias to respond with /dAk/ which could be interpreted as the detection of a fast formant transition even when none was present in the speech signal. This is remarkable in light of the fact that the intensive FastForWord training program is supposed to ameliorate reading problems using such a double manipulation by making speech sounds more discriminable (Merzenich et al., 1996; Tallal, 2004; Tallal et al., 1996). On the basis of these results, I therefore tentatively conclude that there is no difference in the accuracy of the speech perception of the kindergarten children with a familial risk of developmental dyslexia versus no such risk. From a component dominant view, employing a decompositional strategy, I would hence conclude that speech perception, as a possible component process is not impaired!

The dispersion of response times, however, yielded quite different results. Although the mean response times were found to be the same across different stimulus types in both experiments, they clearly differed between groups, with the Familial-Risk group generally showing slower responding than the No-Risk group, although the effect sizes were fairly small (identification: .21; discrimination: .14). The stability of the children's early speech perception was next examined in relation to their later reading performance by correlating the standard deviations for their kindergarten response times to their first-grade reading performance. As expected, the identification task was easier and — probably due to the relatively light memory load — yielded more stable responses than the discrimination task and no significant correlations with the children's later reading performance were found. The standard deviations from the discrimination task, however, showed strongly negative correlations for the unmanipulated and amplified stimuli with the accuracy of performance on the word reading and pseudo-word reading tests for particularly the children with a familial risk of dyslexia. No significant correlations were found for the group of children with no such familial risk. This set of findings suggests that instability of responding to certain types of speech stimuli in kindergarten is systematically related to impaired reading performance in first grade. Those children within the Familial-Risk group with more stable response times as indicated by a lower response time standard deviation performed better on the reading fluency tests in first grade. No differences between the groups of children in the correlations (which were not significant) of their reading performance with the kindergarten measures of correct responding, discrimination or response bias were found.

The question which now arises is why the systematic association between speech perception and reading performance disappears in the context of speech stimuli that have simply been slowed or undergone a double manipulation. One explanation is that these manipulations reduce the constraints on processing and thereby produce more stable performance on the part of the children at risk for dyslexia; the rate of the frequency changes in the speech signal is slowed considerably. This explanation is insufficient as the stimuli that underwent both manipulations were also responded to less accurately by both groups of children. I will address this finding by discussing the results from the analysis of distributions of the response time dispersion based on reading performance.

The results of the reading performance tests in first grade showed 9 out of the 19 children initially at risk for dyslexia to indeed have serious reading problems. This is 47%, which is somewhat higher but consistent with other studies of familial risk for dyslexia. Elbro and Børstrom (1997), for instance, reported that 40% of their much larger sample turned out to be dyslexic. The nine children found to have dyslexia also generally showed a larger dispersion of response times than the group of 29 children with no reading problems in first grade; their kindergarten performance on the speech stimuli which were both slowed and amplified constituted an exception to this pattern, however, and were more uniform with respect to the stability of the response times. The analyses showed relatively more errors to be made on these stimuli by all of the kindergarten children but also revealed a larger response bias (i.e., tendency to respond with /dAk/ when the stimulus was /bAk/) and a tendency to respond "different" when there was, in fact, no difference between the stimuli. When viewed in terms of added constraints to the ability to generate stable performance, one could argue that the response options available to the children were possibly reduced by the double manipulation to a single response option, namely /dAk/ or "different." This would explain the lack of a correlation between the reading performance of both groups of children and the standard deviations for their perception of these stimuli: Their responding was stable, due to the reduced number of options, but nevertheless erroneous. This pattern of responding occurred independent of whether the children showed later reading problems or not, which clearly suggests that impaired performance on a task need not entail impaired underlying components or cognitive processes such as phonological representations in this case.

When the dispersion of the response times were examined to the stimuli that were only slowed, the differences between the children with later reading problems and those without were found to be the largest compared to the other stimuli. In other words, the lack of a correlation with the reading tests for the Familial Risk versus No Risk division appears to have a different cause in this case when compared to the double manipulation. While the means for correct responding and the discrimination index A' for the slowed stimuli did not differ significantly from the means of the other stimuli, those means for the slowed stimuli were nevertheless highest for both groups in both the identification and discrimination tasks. Thus, I have shown a complex interplay of influences of measurement contexts on the observed emergence of impaired performance in speech perception and the observation of associations between impaired-speech perception and reading fluency. In terms of causation I interpret these results as originating from a system in which interactions between components dominate behaviour, not the components themselves.

2.6.1 Context Sensitivity = Stability

To summarise, two new insights have been provided into the relations between speech perception and impaired reading. First, although the accuracy of the kindergarten speech perception of later dyslexic readers did not differ from that of later average readers, the stability of kindergarten speech perception as expressed by the dispersion of response times was strongly and significantly related to reading performance one year later. The higher dispersion of response times associated with impaired reading performance is taken to indicate an unstable system. For proficient reading performance, no relation to the dispersion of kindergarten response times was found. Future research should certainly pursue the possibility of observing such a relation for other deficits such as the visual or motor magnocellular deficits as well. Second, a microanalysis of the speech perception results showed that the speech manipulations within the brief trial time frame of the identification and discrimination tasks imposed additional constraints upon the children's speech perception and thereby allow impaired performance to emerge for both no-risk and familial-risk groups.

The findings presented here involve a dissociation of kinds between accuracy of performance (and therefore also impaired performance) and stability of performance which is gauged by specific constraints interacting on different levels of analysis and measurement context: The acoustic manipulation of the stimulus, the task demands and the genetic and environmental background of the participants. Deliberately adopted, was a speech-perception task that would elicit considerable correct responding on the part of kindergarten children: To test the hypothesis that the differences in accuracy of speech perception found for various types of tasks in previous research can be explained by measurement contexts which may allow more stable performance to emerge at times (e.g. Ramus, 2004). When the problems identified earlier, with finding a single causal chain of impaired components that always leads to impaired or dyslexic reading, the results presented here suggest such an attempt may be difficult at best and futile at worst. A concise task analysis is usually not conducted and measures of central tendency are the main explanatory variables used in analyses. This is why I choose context relativity instead of context sensitivity: It is not the case that there is a component causing an effect that may be amplified or weakened by a context, the effect itself is the result of the context and should thus be interpreted in relation to this context. Adopting an interaction dominant view of cognitive performance that focuses on intra-individual variation rather than between group variation may loosen the almost preformistic claims found in contemporary literature about biological causes of impaired behaviour, particularly when developmental processes such as literacy acquisition are considered (Molenaar, 2008; Molenaar & Campbell, 2009).

It is beyond the scope of this study to provide a definite answer to my interpretation of the data on developmental dyslexia. In any case, some fundamental doubts about the causal relations hypothesised in component-dominant interpretations of the data on developmental dyslexia have been raised. The utility of an interaction-dominant interpretation of the data on developmental dyslexia and (impaired) cognitive processes in general has been demonstrated in this study.

Notes **Chapter 3: Principled Simulation**

- This chapter may be cited as:

  Hasselman, F. (2014). When opposites attract, repel and deceive: Using Recurrent Neural Computation to Model Multi-stable States. In *Beyond the boundary. An analysis of verisimilitude and causal ontology of scientific claims: Ætiologies of developmental dyslexia as a case in point*. fig**share**. http://dx.doi.org/10.6084/m9.figshare.928518

- The supplemental materials to this chapter may be cited as:

  Hasselman, F., (2014). Beyond The Boundary - 3rd Chapter: Supplemental Materials. Retrieved from Open Science Framework, https://osf.io/g396c

Chapter 3

When Opposites Attract, Repel and Deceive

Using Recurrent Neural Computation to Model Multi-stable States

To understand the dynamics between speed and accuracy of a classification performance in stable and unstable systems suggested by the previous chapter, this chapter explores the attractor dynamics of a recurrent neural network (Hopfield) between two opposing basins of attraction. These attractors represent the classification of a stimulus as either /bAk/ or /dAk/ based on exemplar values along two arbitrary dimensions, here representing features of the speech signal (F2 salience and F2 rate of frequency change). The accuracy of performance is assessed by looking at the classification response after presenting the network with stimuli that vary along the two dimensions. The stability of network performance is examined by its energy state. System instability is introduced as (but not limited to) the presence of a third, weak basin of attraction that will never receive a categorisation response. In fact, what is simulated is the constraining of the degrees of freedom available to the network to arrive at a stable and accurate solution.

3.1 Energy State & Classification Accuracy in Recurrent Neural Networks

The theoretical framework for understanding speech perception performance in the previous chapter assumes an unstable complex dynamical system as an explanation for the behaviour of dyslexic readers. Instability was revealed by the dispersion of the children's response times in a speech discrimination task, but not by the accuracy of their phonemic categorisation or their speech identification. Several suggestions for a cause of this instability can be made based on the available literature: It may be a general instability caused by one of the many observed structural cortical anomalies (e.g. Eckert, 2004), leading to dynamical instabilities (Schöner & Dineva, 2007). It may also be due to the presence of a rudimentary phoneme category or allophone (Serniclaes, 2004, 2006). The allophone which supposedly leads to unstable response times on a discrimination task in this particular example, would lie somewhere along the continuum between /bAk/ and /dAk/. In order to provide additional evidence for my unstable complex dynamical system hypothesis, neural network simulations were conducted. It was not a goal to construct a realistic model of speech perception (e.g. see Hopfield & Brody, 2001).

The simulations are intended, rather, to show how the state space of a dynamic system can become unstable when more degrees of freedom available to the network are constrained (structural anomalies hypothesis) or when an attractor state is added (allophone hypothesis). Technically these two modelling options are identical. One modelling outcome has to be of course that the categorisation performance of the network should not be grossly damaged. The model used in

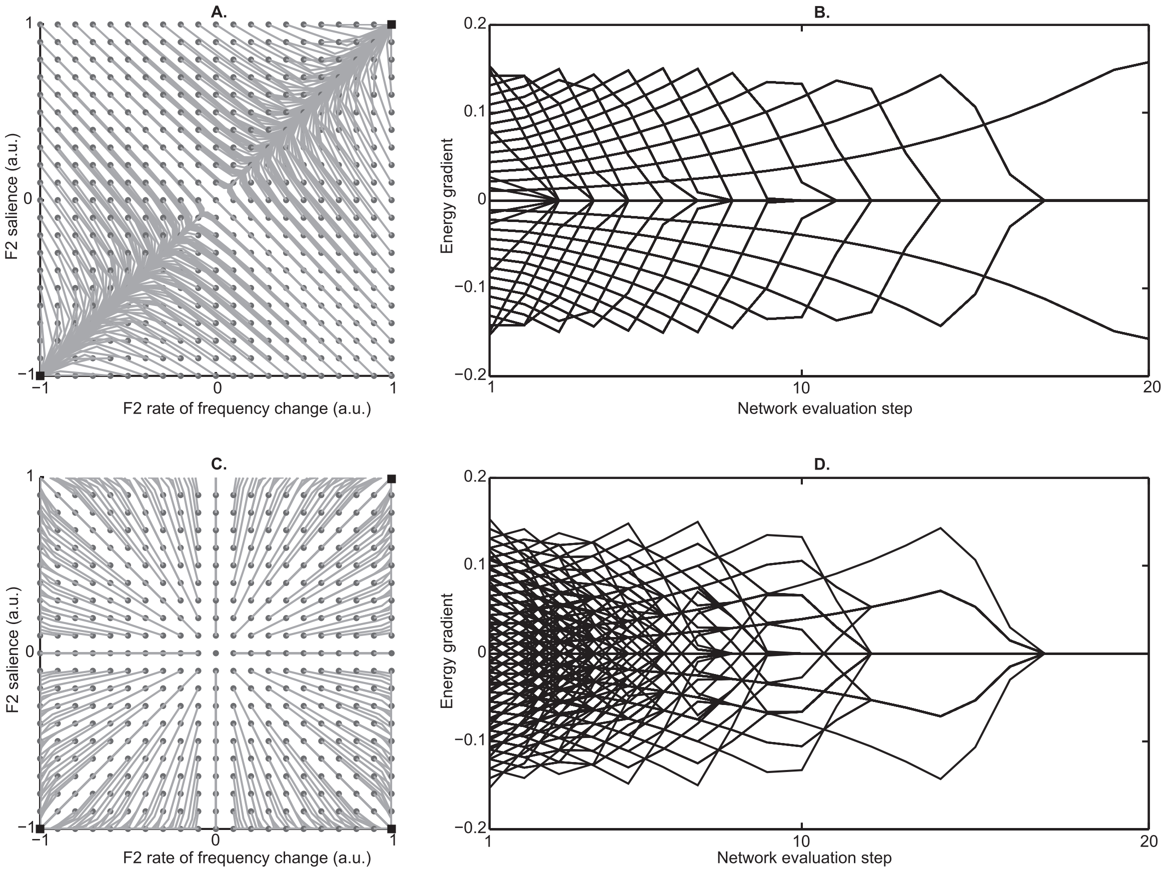


Figure 3.1 – Hypothetical state space spanned by the two parameters manipulated in the speech perception tasks: F2 Salience and F2 rate of change (panels A and C). The state space in panel A has two stable solutions (black squares) and represents average readers. Panel C was created using three stable solutions (black squares) and represents dyslexic readers. The figures show the "routes" through the state space (grey lines) for input varying with respect to the two parameters (grey dots). Panels B and D show the change of network energy for each of the grey dots in panels A and C, respectively.

the simulation is thus intended to characterise the dynamics of the behaviour observed in experiments described in Chapter 2 and not necessarily how the central nervous system processes speech sounds. In previous research, Been and Zwarts (2003, 2004) showed modification of the ARTPHONE model (Grossberg & Schmajuk, 1989) dubbed SWEEP to successfully explain a great deal of dyslexic speech perception data. I consider the level of description used in the model (i.e., brain regions detecting frequency sweeps or formant transitions) untenable, because as the current findings and that of others (see for instance Tuller, Case, Ding, & Kelso, 1994) have shown speech perception to not merely depend upon the detection of specific formant transitions or particular voice onset times (also see chapters 4 and 5). Been and Zwarts (2004, Appendix B) also report that their modifications to the ARTPHONE model made it lose its resonant properties which are the most essential part of Adaptive Resonance Theory (Grossberg, 2003).

3.2 Method and Procedure

3.2.1 Network Architecture: An Example by Principle

The model used was a two-neuron Hopfield network (Hopfield, 1982) as implemented in the MATLAB software (version R2008b, The Mathworks, Inc.), which uses a design method based on linear differential equations defined on the edges of a closed hypercube (Li, Michel, & Porod, 1989). The

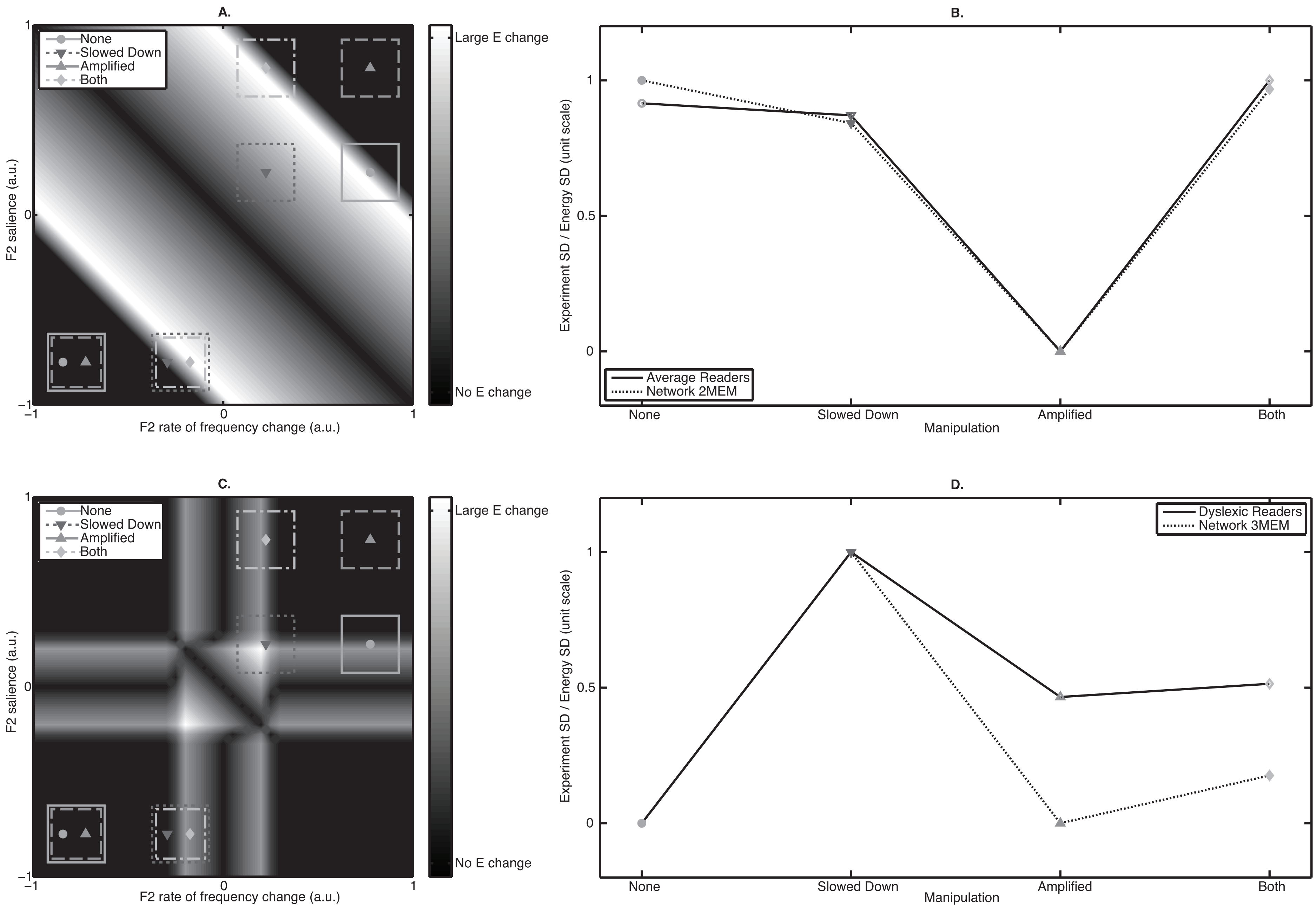


Figure 3.2 – Figure 3.2. Results of 2MEM simulations (panels A and B) and 3MEM simulations (panels C and D) of response time SDs by means of evaluation of network energy after the network processing was stopped. See text for details.

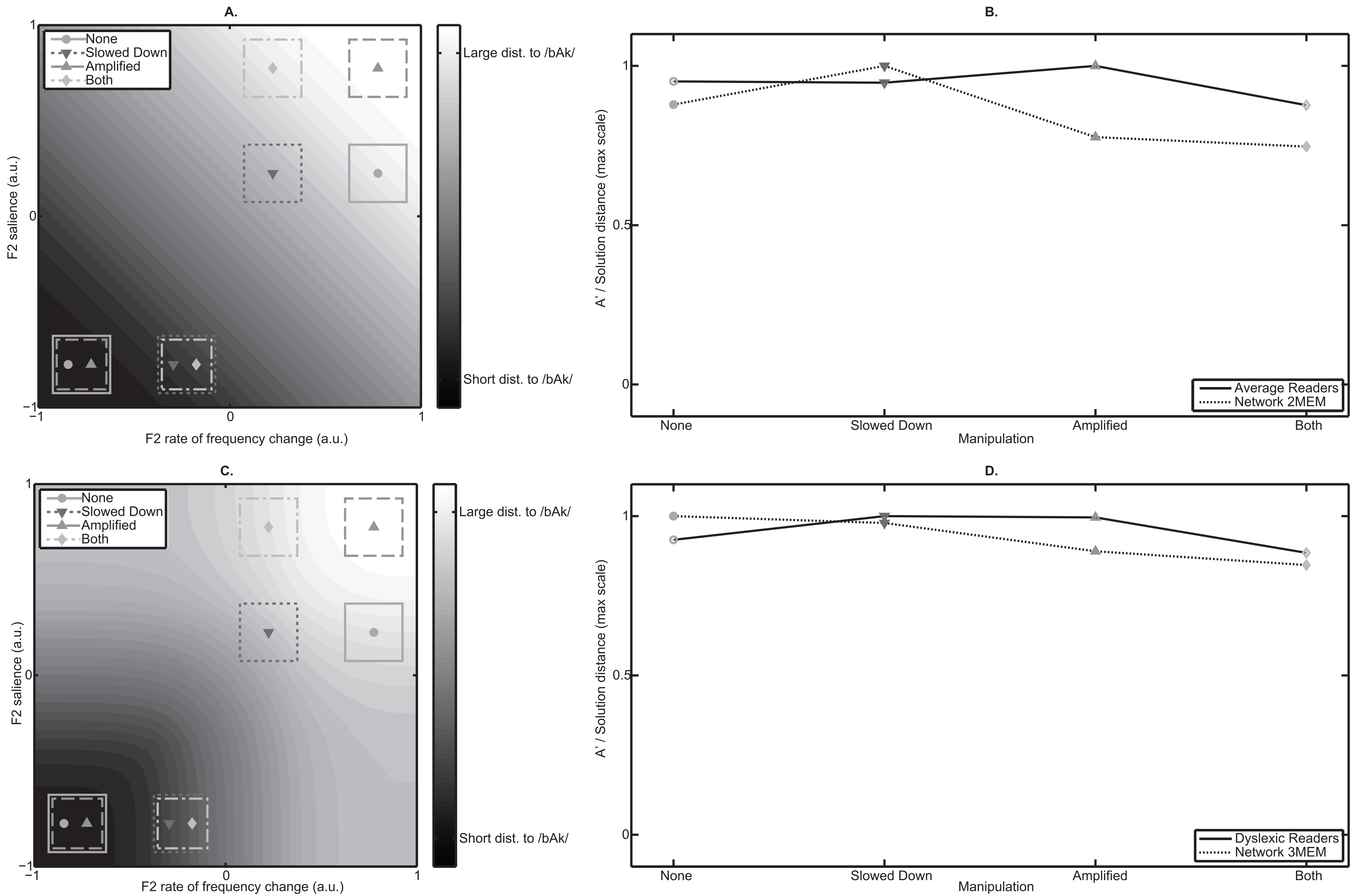


Figure 3.3 – Results of the 2MEM simulations (panels A and B) and 3MEM simulations (panels C and D) of A' by means of evaluation of the distance from the memory location which the network settles upon after being stopped (Only distance to /bAk/ shown). See text for details.

edges of the hypercube define the state space of the system. Hopfield networks are recurrent networks in the sense that network output is fed back as input obeying Hebb's law of associative memory. The purpose of the network is to store information in memory as stable points and thereby allow lightly distorted input to be recognised and categorised as a variation on that information previously stored in memory. This allows creating a state space defined by the two parameters manipulated in the experiments in Chapter 2, being the salience of the second formant transition and the rate of frequency change of the second formant transition. Of course, for the model these interpretations are arbitrary and may be anything.

In Figure **3.1** Panel A, a graphic depiction of the hypothetical state space is shown. The squares in the bottom left corner and top right corner show the locations of the stored dynamic memory representing /bAk/-like and /dAk/-like stimuli, respectively. The grey points show input to the network, which varies with respect to the two parameters. The grey lines depict the path the network chooses to categorise such input. Clearly two basins of attraction can be distinguished and seem to divide the state space into two equally large parts. The oblique line in the centre represents a special series of points that fall exactly between the two attracting forces and may thus cause them to be categorised at a location, which is not a predefined memory state. These special states are called spurious states. In dynamic systems terms, this line is called a repellor (i.e., the opposite of an attractor) and, within the context of speech perception, it can be associated with a phoneme boundary. A very interesting property of the network is that its state at each evaluation step can be described by an energy function that belongs to a class of functions called Lyapunov functions. Such functions "seek" stable states by minimising their value according to a dynamic rule. This can be seen in Figure **3.1**, Panel B, which is the plotted change in energy for each point in panel A. Most points reach a stable state before 20 evaluation steps when the energy change equals zero.

3.3 Simulation of Results

Using the concepts of attractors and energy states, I can attempt to find parallels to the proposed framework. Figure **3.1**, Panel C, shows a state space to which an extra memory location has been added in the lower right corner (dubbed: 3MEM). This could represent the natural allophone category that has been assumed to still exist in dyslexic readers. It could also and equally likely imply that not enough degrees of freedom are available as recurrent connections between the neurons, to yield a three-category solution. The attracting forces within the state space have changed as a result of the addition of this extra memory location and additional repellor, or in the context of speech perception; a phoneme boundary has emerged (represented by the horizontal line at 0). The spurious points that existed in the network with only two memory locations (2MEM) have been given greater attractive force. The exact location of the third memory point for such a pattern to emerge is not of particular importance provided it is proximal to at least one of the spurious points that existed in the 2MEM network. Serniclaes et al. (2004) argue that the allophone category is, in principle, present in average readers but not as strongly as in dyslexic readers.

The energy state has also been altered as can be seen in Figure **3.1**, panel D; there are far more unique initial energy states compared to the 2MEM network of Figure **3.1**, Panel B. In this view, the 3MEM network is shown to be less stable than the 2MEM network. It should be noted, however, that the two networks are basically the same system, and obey the same dynamic rules of associative memory. I have just added an extra response option, which is in keeping with how other authors have handled the impaired stability/plasticity problem (Colangelo, Holden, Buchanan, & Van Orden, 2004; Farrar & Van Orden, 2001). That is, response options that are basically not available during unimpaired performance may be available during impaired performance. Stated differently, in cases of impaired performance, there are not enough constraints to make the system converge upon a single stable state. In recurrent neural networks, the connection weights between neurons are viewed as constraints upon the state space for the system. In a dynamic process, the weights change to bring the network into a stable state and effectively decrease the degrees of freedom that characterise the system. So what is actually manipulated here are the degrees of freedom available to the network by adding another memory state. Again, the interpretation is quite arbitrary, what is important are the dynamics being modelled.

It must also be assumed that the state space and hence the parameter values corresponding to the speech stimuli in the model being simulated is completely defined by the stimuli themselves. It is not assumed that the memory locations within the network represent static, symbolic representations of speech sounds but, rather, that the memory locations are the dynamic representations of the poles of an artificial continuum created by the tasks in the study described in Chapter 2. When the salience of F2 is considered then, the highest value corresponds to the amplified stimuli. Stimuli not amplified but with a formant transition (i.e., the slowed version of /dAk/ and the unmanipulated /dAk/ stimulus) receive a lower value for this salience parameter. And lowest are the unmanipulated /bAk/ stimulus and the slowed version of /bAk/. The parameter rate of F2 frequency change has also been used in the simulations by Been and Zwarts (2003, 2004). The rate of change becomes smaller when a stimulus is slowed because the frequency decreases across a larger time frame. To simulate this, those stimuli that were slowed down were placed closer to each other.

3.3.1 Simulation of Response Time SD differences

An energy change of 25 points was next sampled for each stimulus type from the regions marked in Panels A and C of Figure **3.2** (the contours show the values of energy change). The same points were used for both networks. For each type of stimulus, the mean of these points was calculated. In light of the fact that a categorisation task was being simulated, the difference between the /bAk/ and /dAk/ stimuli of these averages for each stimulus type was taken as a measure corresponding to the dispersion of response times (Energy SD). In order to compare their relative value, all values were transformed to the unit scale by subtracting the minimum value and dividing the result by the maximum value for the result. The results are shown in Panels B and C of Figure **3.2**. For the 2MEM network, the order is almost identical to the actual SD data. For the 3MEM network, the amplified stimuli reached the same value as the unamplified stimuli, which was the minimum and thus 0. The relative order for the amplified stimuli and stimuli with both manipulations is preserved, however.

3.3.2 Simulation of Discrimination Performance (A’)

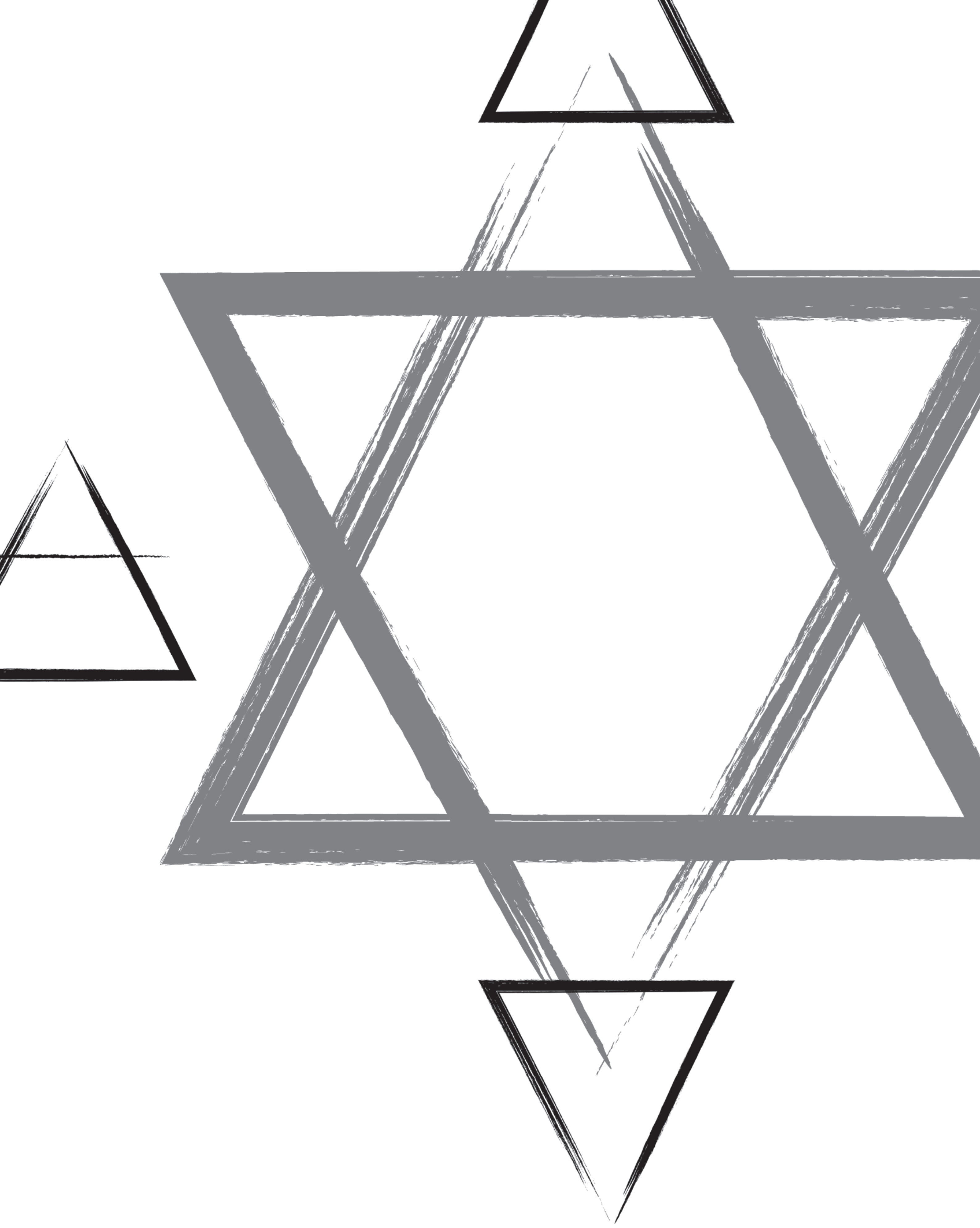
A straightforward measure of discriminability for the points defined representing the stimuli is the distance of a point from a stable memory location at a particular evaluation step. The same 25 points per stimulus type were sampled as in the SD simulation and stopped the 2MEM network again at 5 evaluation steps and the 3MEM network again at 10 evaluation steps (Figure **3.3**, panels A and C). For /bAk/-like stimuli, the distance to the memory in the lower left corner location was calculated. These distance values are shown as the contours in Figure **3.3**; for the /dAk/-like stimuli, the distance to the memory location in the upper right corner was calculated (values not shown inFigure **3.3**). Then simply the average of the /bAk/-like stimuli and the /dAk/-like stimuli was taken for each stimulus type. A value close to zero meant good discriminability because the /bAk/-like stimulus was close to the lower left memory state and the /dAk/-like stimulus was close to the upper right memory state. Given that the direction of this measure is the opposite of the discrimination index A’ with higher values associated with greater discriminability, it was decided to invert the values obtained from the network. Given that the A’ values did not differ much within the different groups of readers per stimulus type, comparison of the relative orders for the stimuli was not informative. In order to examine the possible differences between the simulated and observed data in terms of the magnitude of discriminability, max scaling was applied (i.e., division by the maximum value in the data).

Both sets of data now have a maximum value of 1 with all other points deviating in magnitude relative to this value. Panels B and C in Figure **3.3** show the results with the A' data coming from the 29 average readers and 9 dyslexic readers. The mean solution distance does not vary greatly across the stimuli, as expected.

3.4 Conclusion and Discussion

The results from the network simulations show it is possible to understand these experimental results in terms of the state space of a dynamic system with enough constraints to provide an accurate categorisation of the different stimuli, while the stability of the categorisation does differ per stimulus type as observed in the empirical results. More importantly, different patterns of stability can be seen for the two networks, which show that the addition of an extra response category in the case of the 3MEM network could well be a plausible cause for the destabilisation in dyslexic readers' speech perception.

It is important to note that the 3MEM network has fewer constraints relative to the 2MEM network in terms of network connections available to decide which category should be chosen and is therefore less stable. It is entirely possible that these limited constraints are the actual mechanism driving both this simulation results and Serniclaes et al.'s (2004) findings instead of the actual existence of a rudimentary phoneme category (i.e., the dynamic instability may have a more general origin as proposed by Schöner & Dineva, 2007). This notion is more in line with what other authors have suggested about the dynamic emergence of categories in speech perception (Tuller et al., 1994).



In the most modern theories of physics **probability** seems to have replaced aether as the nominative of the verb 'to undulate'.

- Sir Arthur Stanley Eddington (1935 New Pathways in Science, p.110).

The classical elements have an alchemical form (triangular). The quintessence in alchemy was the Æther, often discarded and re-introduced as is evident from this quote:

The **aether**: Invented by Isaac Newton, reinvented by James Clerk Maxwell. This is the stuff that fills up the empty space of the universe. Discredited and discarded by Einstein, the aether is now making a Nixonian comeback. It's really the vacuum, but burdened by theoretical, ghostly particles.

- Leon Lederman

Apocrypha

Lederman, L.M., and Teresi, D. (2006). The God Particle: If the Universe is the Answer, What is the Question.

Notes | **Chapter 4: Strong Inference**

- This Chapter has been published as:

  Hasselman (2015), Classifying acoustic signals into phoneme categories: average and dyslexic readers make use of complex dynamical patterns and multifractal scaling properties of the speech signal. *PeerJ 3:e837*; DOI 10.7717/peerj.837

- The on-line version contains meta-data (views, downloads), but also the review history and comments and questions https://peerj.com/articles/837/

- The supplemental materials to this chapter may be cited as:

  Hasselman, F., (2014). Beyond The Boundary - 4th Chapter: Supplemental Materials. Retrieved from Open Science Framework, https://osf.io/2r5eh

Chapter 4

Classifying Acoustic Signals Into Phoneme Categories

Average and dyslexic readers make use of complex dynamical patterns and multifractal scaling properties of the speech signal

In Chapters 2 and 3 the distinction between the accuracy and the stability of performance was introduced and related to the classification and discrimination of two phoneme category exemplars, /bAk/ and /dAk/. As was shown, manipulations of the ratio with which spectral features change over time and/or manipulations of the relative power of dynamically changing features in the spectrum were associated with the stability of the classification performance. This was especially the case for those children who turned out to have standardised reading fluency scores in the lowest 25% of the population one year later. This instability in terms of response times was found to be extremely dependent on the measurement context, that is the particular stimulus type presented in the experiment, whereas classification and discrimination performance remained similar and consistent between and within the two groups of children. In Chapter 3, recurrent neural network simulations showed that this distinction might be understood as the attractor dynamics in a two-dimensional phase space governed by recurrent neural dynamics, when typical exemplars of the phoneme categories are considered. I hypothesised this phase space was spanned by the dimensions rate of F2 change and F2 salience The choice to interpret the two dimensions as frequency characteristics of the speech signal is arbitrary and inspired by the acoustic manipulations based on the auditory temporal processing hypothesis. There are, as stated earlier, many more features of the speech signal that authors believe to be of importance for speech perception in dyslexic readers. Nothing in the architecture of the model prefers one interpretation to another; the dynamics in the phase space stay the same irrespective of interpretation. The network architecture thus very much represents a theoretical structure as implied in Chapter 1.

This Chapter will examine three types of measures that represent different distinguishing features of the speech signal and attempt to uncover which features are actually used by participants to classify the stimuli into speech sounds. I will not construct stimuli that exclusively represent these measures as is common in speech and auditory perception studies (see e.g. Boets, Ghesquière, van Wieringen, & Wouters, 2007; Pasquini, Corriveau, & Goswami, 2007). Instead, I will extract all measures from one and the same set of stimuli and analyse which measures enable a classification response (by a simple classifier algorithm) that is most similar to that observed in the participants. The text is taken from the article published in PeerJ (See Chapter Notes).

4.1 Abstract

Several competing aetiologies of developmental dyslexia suggest the problems with acquiring literacy skills are causally entailed by low-level auditory and/or speech perception processes. The purpose of this study is to evaluate the diverging claims about the specific deficient perceptual processes under conditions of strong inference. Theoretically relevant acoustic features were extracted from a set of artificial speech stimuli that lie on a /bAk/-/dAk/ continuum. The features were tested on their ability to enable a simple classifier (Quadratic Discriminant Analysis) to reproduce the observed classification performance of average and dyslexic readers in a speech perception experiment. The 'classical' features examined were based on component process accounts of developmental dyslexia such as the supposed deficit in Envelope Rise Time detection and the deficit in the detection of rapid changes in the distribution of energy in the frequency spectrum (formant transitions). Studies examining these temporal processing deficit hypotheses do not employ measures that quantify the temporal dynamics of stimuli. It is shown that measures based on quantification of the dynamics of complex, interaction-dominant systems (Recurrence Quantification Analysis and the multifractal spectrum) enable QDA to classify the stimuli almost identically as observed in dyslexic and average reading participants. It seems unlikely that participants used any of the features that are traditionally associated with accounts of (impaired) speech perception. The nature of the variables quantifying the temporal dynamics of the speech stimuli imply that that classification of speech stimuli cannot be regarded as a linear aggregate of component processes that each parse the acoustic signal independent of one another, as is assumed by the 'classical' aetiologies of developmental dyslexia. It is suggested the results imply that the differences in speech perception performance between average and dyslexic readers represent a scaled continuum rather than being caused by a specific deficient component.

4.2 Introduction

Many aetiologies of developmental dyslexia assume some deficit in auditory processing may be causally entailed in the difficulty with acquiring proficient levels of reading and spelling ability experienced by a small percentage of the population (see e.g., Ramus, 2004). The nature of the features of the acoustic signal that are assumed to be able to evidence such deficient components (e.g., phoneme representations, allophones) or component processes (e.g., frequency sweep detection, rise time perception) varies greatly between aetiologies (cf. Serniclaes and Sprenger-Charolles, 2003). The purpose of this study is to compare a number of such features under conditions of strong inference (cf. Platt, 1964). The goal is to examine whether average and dyslexic readers actually use these features to arrive at a particular classification of a speech stimulus, a first and necessary step to take before their causal entailment in dyslexic reading can be claimed. Three types of measures will be examined that represent different distinguishing features of the speech signal, however, not by constructing stimuli that exclusively represent these measures as is common in auditory and speech perception studies (see e.g., Boets et al., 2007; Pasquini et al., 2007). Instead, all measures will be extracted from one and the same set of stimuli and measures will be evaluated on their ability to enable a simple classifier to yield a response that is similar to classification responses by participants.

The measures used in this study can be extracted from any continuous signal (sampled, synthesised, or generated otherwise), but are very different in the type of information they are thought to capture, or more suitably: represent. The first are *Component Process Measures*, derived from the signal because of their supposed importance in contemporary theoretical assumptions about deficient components of cognitive or sensorimotor processes related to developmental dyslexia and speech perception. They represent the Component Dominant family of dyslexia ontology. The second type of measure are *Periodicity Measures*, derived from (linear) transforms or decompositions of the signal used in other contexts to express the average periodicity, harmonicity or regularity of the 'true' signal (see e.g., Guiard, 1993, for an application to harmonic movements). These measures quantify periodic changes of the variable in question over time. The third are *Complex Dynamic Pattern Measures* derived from nonlinear time series analyses and multi-scale analyses that have a wide range of applications in the general study of the behaviour of complex dynamical systems. The interaction-dominant perspective on explaining complex behaviour assumes it emerges out of the interactions of many processes fluctuating on different spatiotemporal scales. The Complexity Matching or Complexity Control hypothesis posits that humans make use of the invariant structure of such complex dynamical patterns to coordinate their behaviour in ways that are comparable to principles for optimal and maximal information transport between complex systems as posited by formal fluctuation dissipation theorems (e.g., the '1/f resonance hypothesis' Aquino et al., 2011). Complexity science has developed a number of analyses that allow a quantification of complex temporal patterns and self-affine structure in empirical time-series. Such measures often concern a quantification of dynamics in a phase-space representation of the signal, reconstructed by means of delay embedding methods (cf. Kantz and Schreiber, 2003), or, the assessments of scaling relations between signal variability and the temporal resolution at which the variability is assessed (cf. Kantelhardt, 2011). The techniques used in this article to quantify phase-space dynamics and scaling relations in the speech signal are Recurrence Quantification analysis (RQA, cf. Marwan et al., 2007) and Multifractal Detrended Fluctuation Analysis (MF-DFA, see Kantelhardt et al., 2002; Ihlen, 2012).

The latter two types of measure (Periodicity and Complex Dynamic Pattern measures) have not been the focus of studies on dyslexia and speech perception, even though these measures seem tailor made to test claims of deficits in detecting complex dynamic frequency or amplitude patterns present in the speech signal.The association between speech perception and non-linear behavioural phenomena (e.g. hysteresis, enhanced contrasts) has been established in a number of studies (see e.g., Case et al., 1995; Porter and Hogue, 1998; Tuller et al., 1994; van Lieshout et al., 2004; Hasselman, 2014a). Recent studies have shown that quantification of recurrent patterns (RQA) and the presence of power-law scaling in trial series of word-naming latencies of dyslexic readers are different (more random, less fluent) from average readers and are correlated to reading performance on standardised tests. The correlation only appears in dyslexic readers (Wijnants et al., 2012b). A comparison of response latency distributions in different tasks (word-naming, colour-naming, arithmetic, flanker tasks), suggests dyslexic readers' response distributions are a scaled versions of average readers, in which the relatively larger 'heavy tails' account for more variable, more random behaviour (Holden et al., 2014). This would indicate a general scaled continuum account of dyslexia and not, as component dominant aetiology suggests, a localised specific deficit. This is reflected in how the temporal evolution and change processes (i.e. continuous dynamics) are studied: Component process measures quantify change over time as a nominal variable that can be 'on' or 'off' in a stimulus (F2 rate of frequency change is high or low; Rate of change of envelope modulation is high or low). This is not the same as quantifying the dynamics of a continuous signal (RQA), or the full range of temporal correlations present in a signal (multifractal spectrum).

Figure **??** displays six different representations of a single speech stimulus (Stimulus 1) that was used to extract measures that have been suggested to be important for understanding the role of speech perception in the aetiology of developmental dyslexia. Each stimulus representation can be ordered with respect to the component versus interaction dominant causal ontology used in hypotheses about the origins of impaired performance associated with developmental dyslexia. What follows will be an introduction to the different measures used in this study and an analysis of their ability to serve as the features that enable classification of speech stimuli as observed in the performance of average and dyslexic readers in simple labelling experiments of those stimuli.

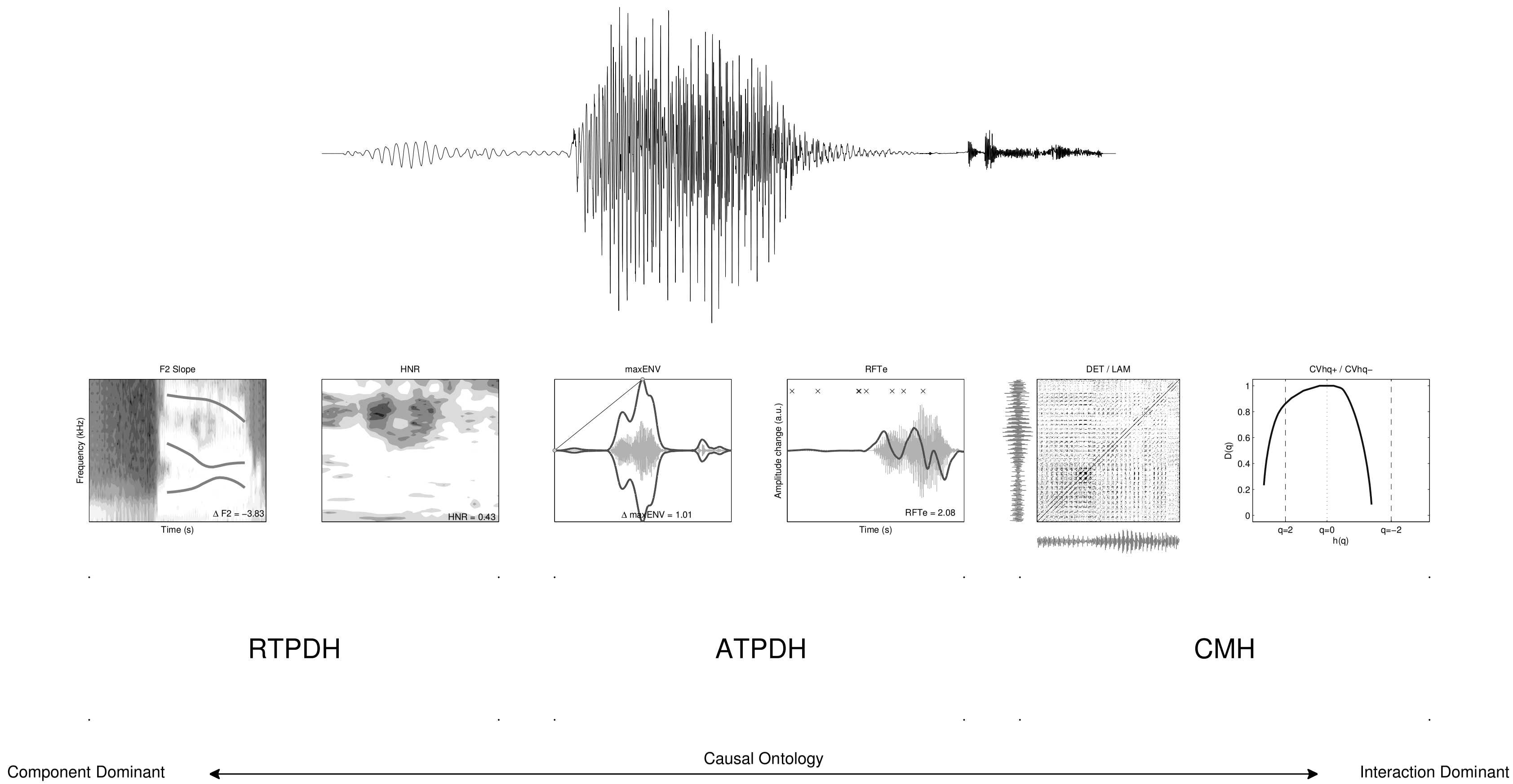


Figure 4.1 – Six representations of stimulus 1 used to extract the features for classification by Quadratic Discriminant Analysis.

4.3 Component Process Measures: What does Temporal refer to?

The "temporal" auditory processing deficit hypotheses concern properties or information content in auditory stimuli that cannot, due to the rate with which the information changes over time, be properly perceived by the person afflicted with the deficit. There are two major deficit hypotheses of this kind: The auditory temporal processing deficit hypothesis (ATPDH: Farmer and Klein, 1995; Tallal, 2004) and the rise time perception deficit hypothesis (RTPDH) proposed by Goswami and colleagues (see e.g., Goswami et al., 2010a, 2002)

The ATPDH states that speech stimuli with rapid transient spectral elements are processed less accurately because such elements occur too fast to be perceived by people with the processing impairment. In fact, the claim is not limited to spectral features, but pertains to any sequence of auditory stimuli presented in rapid succession.

Tests that have been employed to reveal this deficit are for instance temporal order judgements (e.g., Pasquini et al., 2007) and auditory gap (or threshold) detection (Boets et al., 2007; Corriveau et al., 2007). There is also evidence from neuroscience that seems to point to anomalous functional responses to rapid auditory stimuli (Temple et al., 2000) or an "asynchrony" in the speed of processing between auditory and visual modalities (Breznitz, 2003). Note that essentially, these are two different deficits:

1. An auditory stimulus with rapidly changing elements is not detected / processed adequately.
2. The speed with which processing of auditory stimuli takes place is not adequate (out of sync).

From the literature it is unclear which of these two temporal deficits the ATPDH actually refers to, in fact both can be true at the same time. The early work by Tallal and co-workers suggests the first option (see e.g., Tallal, 1976; Tallal et al., 1993; Tallal and Piercy, 1974). However, since the ATPDH has been "adopted" by the magnocellular theory of dyslexia (Stein, 2001; Stein and Walsh, 1997), option seems more appropriate. This magnocellular theory states that the sensorimotor deficits observed in dyslexic readers may be explained by the anomalies found in the magnocellular neural pathways responsible for fast information transferral. It is thus not exactly clear what the "temporal" in temporal processing refers to. A similar problem plays a role in the rise time perception deficit (e.g., Livingstone et al., 1991).

The RTPDH states that there are problems with the perception of the slow changing amplitude modulation cues, or rise times of the amplitude envelope of the speech signal. Temporal here thus refers to the opposite of ATPDH in terms of the rate of change involved.

The hypothesis has recently been placed in a temporal sampling framework (Goswami, 2011) that provides a neurocognitive basis for the deficit. The main explanatory work in the theory is done by the fact that perceiving changes in amplitude envelopes is essential for segmenting the speech stream into smaller units, for perceiving prosody to mark boundaries of sentences, words and syllables (Ziegler and Goswami, 2005). In one of the first publications presenting this hypothesis (Goswami et al., 2002), it is suggested that the deficit concerns the processing of the acoustic structure of the syllable, which is best described as rhythm detection. This was tested by asking children to distinguish between stimuli on a continuum from smaller (15 ms) to larger (300 ms) envelope rise times of the modulating wave. The slope of the psychometric categorisation function of the dyslexic readers was smaller than that of typically developing children (compared to chronological age and reading age). The conclusion was that the dyslexic readers were not detecting the envelope onsets that make up the beat of the signal. Performance on the envelope onset detection task explained more variance in reading and spelling performance than the temporal order judgement tasks and rapid frequency discrimination tasks associated with ATPDH. This deficit is also thought to have broader consequences for meter and beat perception in music by dyslexic readers (Goswami, 2006; Huss et al., 2010). It is suggested that a deficit in beat perception may also explain why dyslexic readers have problems producing speech, or tapping to a metronome (Corriveau and Goswami, 2009). The causal connection to reading is however still through a deficient representation of a phoneme-like structure due to poor beat perception. This is why the hypothesis belongs in the arena of the component dominant ontology.

The question remains, what exactly is the process that is deficient here? The authors (Goswami et al., 2002; Goswami, 2006; Goswami et al., 2010a; Goswami, 2011) use “rise time perception deficit”, “envelope amplitude onset detection deficit”, “perceptual insensitivity to amplitude modulation”, “beat perception deficit” and “p-centre detection deficit”. Recently, the perception of fast spectral changes in formants was directly compared to rise time perception in a /bA/-/wA/ continuum on which stimuli differed either by frequency rate of change or envelope rate of change (Goswami et al., 2010a). The frequency onsets of the formants were kept equal in both conditions. It was concluded that dyslexic children were poor at discriminating between sounds based on the rate of change of the envelope, whereas discrimination based on formant transition duration (the rate of change of frequency) yielded normal performance. The authors interpreted the results as a failure to detect envelope cues by dyslexic readers, not rapid frequency changes. What does this imply? Are there too many, or too few rise time onsets in the signal to be perceived. Or, if the deficit is indeed also responsible for anomalous rhythm production, is it a matter of a deficient coupling between an internal clock and an externally perceived rhythm as suggested in the temporal sampling framework (Goswami, 2011)? If those rise time onsets were made more salient, would they lead to better beat perception? Is it a deficit in perceiving the rate with which the amplitude envelope changes in the signal instead of the actual detection of the onset of the envelope? This is what is suggested by the stimuli used in (Goswami et al., 2010a) and seems a different, more specific auditory processing deficit than the more general deficit the same authors proposed to detect the occurrence of envelope onsets as a beat or rhythm.

Confusion about the specifics of the characteristics of the stimuli to which the deficits pertain seems to occur in both hypotheses: Periodicity or pattern detection versus rate of change detection. The measures that will be extracted from the speech signal in this study will address both features of the complex speech signal. The measures that seem to relate most to a deficient component process appear to be the rate of change of the formant frequency and the rate of change of the amplitude envelope. The periodicity, or pattern measures will be discussed in the next paragraph. To obtain the rate of change of the formant frequency of a stimulus, the Fourier transform of the speech signal is taken and formant tracks are extracted from the spectrum. The slope of the second formant (F2) in the spectrogram is calculated as a measure of rate of frequency change. For RTPDH there are several options to quantify the rate of change of the amplitude envelope. Here, the stimuli used in (Goswami et al., 2010b) are considered an appropriate measure, because in that study several of the options mentioned above (rise time duration, envelope onset, tempo, etc.) are contrasted against one another. Differences between dyslexic readers and typical readers (chronological age controls) in that study were significant when discriminating between two types of stimuli: a) Stimuli with single ramp envelope onsets (with random steady states and rise times varying from 15 to 300ms); and b) composite stimuli consisting of a standard rise time (15 ms) alternated with a longer rise time (up to 192 ms). The study showed that performance on discrimination tasks with these stimuli was correlated with rhyme detection and reading and had a unique contribution to explained variance in these variables in a regression model. A sensible measure then seems to be the time it takes for the amplitude envelope to rise to its maximal value, which also marks the onset of the rhyme (b-Ak). To obtain the measure, first the absolute value of the Hilbert transform of the signal is taken (c.f. Feldman, 2008; Smith et al., 2002), which yields the immediate envelope. The slope of the line one can draw from the amplitude envelope at the start of the signal to its maximum value is considered an estimate of the most important slow rise time that needs to be detected in order to distinguish between speech stimuli.

Acoustic manipulations of the speech signal, based on the ATPDH refer to amplification or slowing down (or both) of the fast spectral changes present in the speech signal. These manipulations are expected to also affect the amplitude envelope, which is important for RTPDH. Amplification may lead to steeper rise time slopes whereas slowing down the signal is expected to lead to (relatively) slower rise times. (Been and Zwarts, 2003) presented simulations of the effect of amplification of the fast formant transitions using their SWEEP model. The SWEEP model is a dynamical model built around the assumption that speech perception involves detection of frequency sweeps. They predicted that the amplification manipulation would indeed lead to a better performance on behalf of the dyslexic readers. Following this line of reasoning, we may expect a measure that indexes the rate of change of a formant transition in a speech signal to be a measure of which ATPDH would agree dyslexic readers cannot maximally exploit to identify and discriminate between speech sounds.

In Figure **4.2** the spectrograms of the stimuli used in the present study are plotted. The rate of change of the formant transition calculated as the slope of F2 in the spectrum is given for each of the 40 stimuli. Figure **4.3** shows the smoothed amplitude envelopes of all the stimuli and the rise time is calculated as the slope from the start of the stimulus to the maximum amplitude. As shown in the figure, these measures differ between the stimuli and are thus candidate features that may actually be used by participants.

4.4 Periodicity Measures: Harmony of Frequency and Amplitude

The periodicity measures used are Rise- and Fall-Time Entropy (RFTe) and Inharmonicity (also known as Harmonics-to-Noise-Ratio, HNR). In theory these measures should be connected to RTPDH and ATPDH respectively. Quite remarkably, to my knowledge they have never been used in studies in the context of speech perception and developmental dyslexia. RFTe represents the entropy (disorder) in the distribution of rise and fall times estimated present in the envelope. It is calculated by taking the first derivative of the immediate amplitude envelope (obtained by taking the absolute value of the Hilbert transform of the signal), which represents the rate of change of the amplitude. When the differenced amplitude envelope changes sign, that is crosses the x-axis, there is a peak in the amplitude (rate of change is zero) after which the amplitude rises or falls. Quantifying the time between peaks in the envelope by subtracting the time stamp of subsequent zero-crossings in the derivative thus yields a distribution of durations; the time it takes for the amplitude to rise or fall. The entropy of this distribution of discrete durations of size n can be calculated as the chance of observing a particular rise or fall time $pRFT_i$ (equation 4.1) and inserting it into the regular formula for Shannon entropy (equation 4.2).

$$pRFT_i\& = \frac{RTF_i}{\sum_{i=1}^{n} RFT_i} \tag{4.1}$$

$$RFTe = -\sum_{i=1}^{n} pRFT_i * \log_2(pRFT_i) \tag{4.2}$$

RTFe may be considered an estimate of the harmony of the perceptual rhythm invoked by amplitude changes. High entropy means that there is disorder or noisiness in the amplitude envelope of the signal. Another way to interpret entropy is in terms of information: The value of the entropy denotes how many bits of information (due to $\log_2$) would be needed to predict the rate of change of the envelope. More bits needed means less regularity and more disorder in the curve. The RTFe

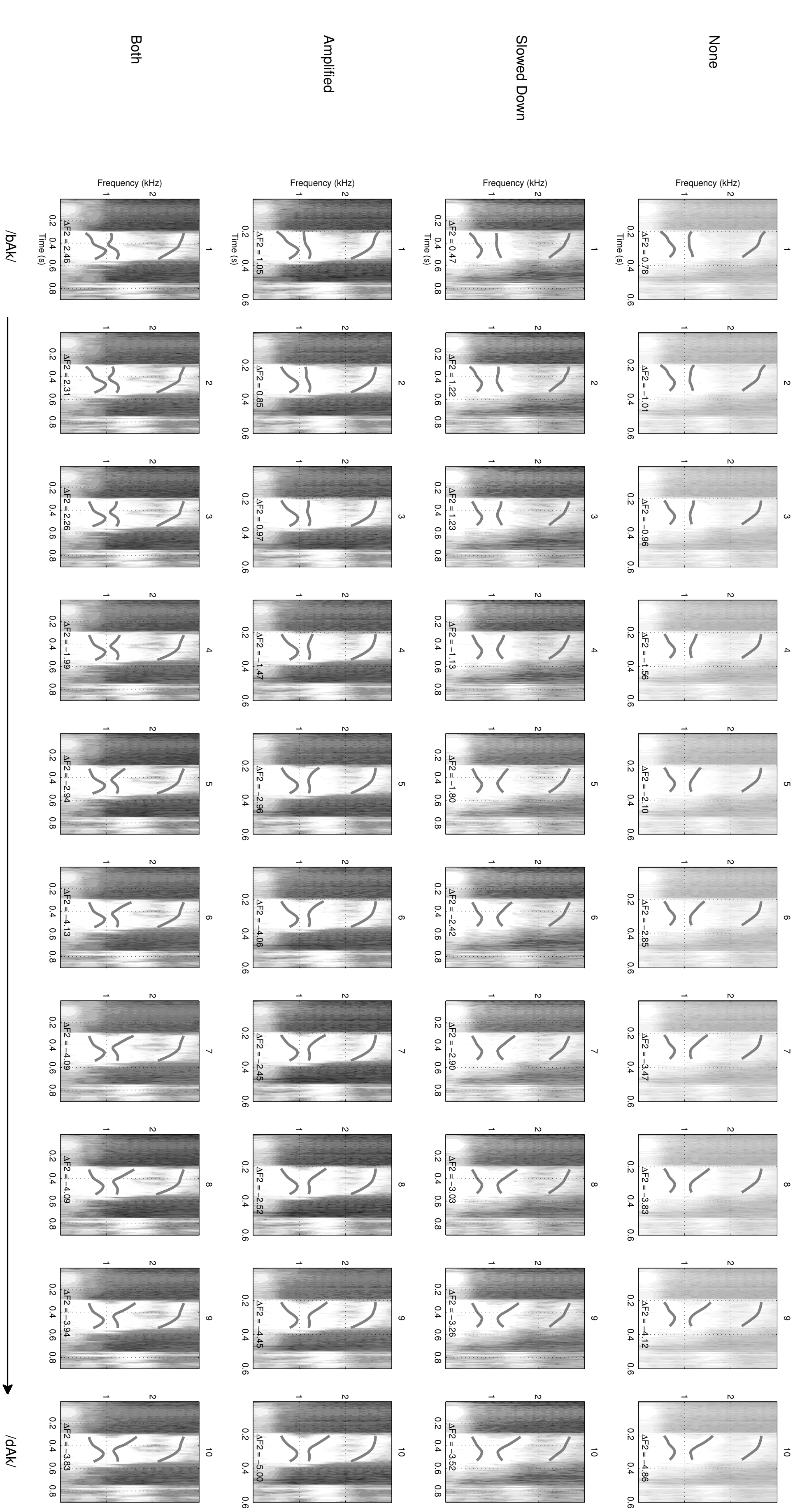


Figure 4.2 – Figures represent spectrograms of the stimuli used in the experiment. There are four manipulations of the 10-step continuum. Formant Sweeps (△ F2) are calculated as the slope of the second formant transition (the second white line).

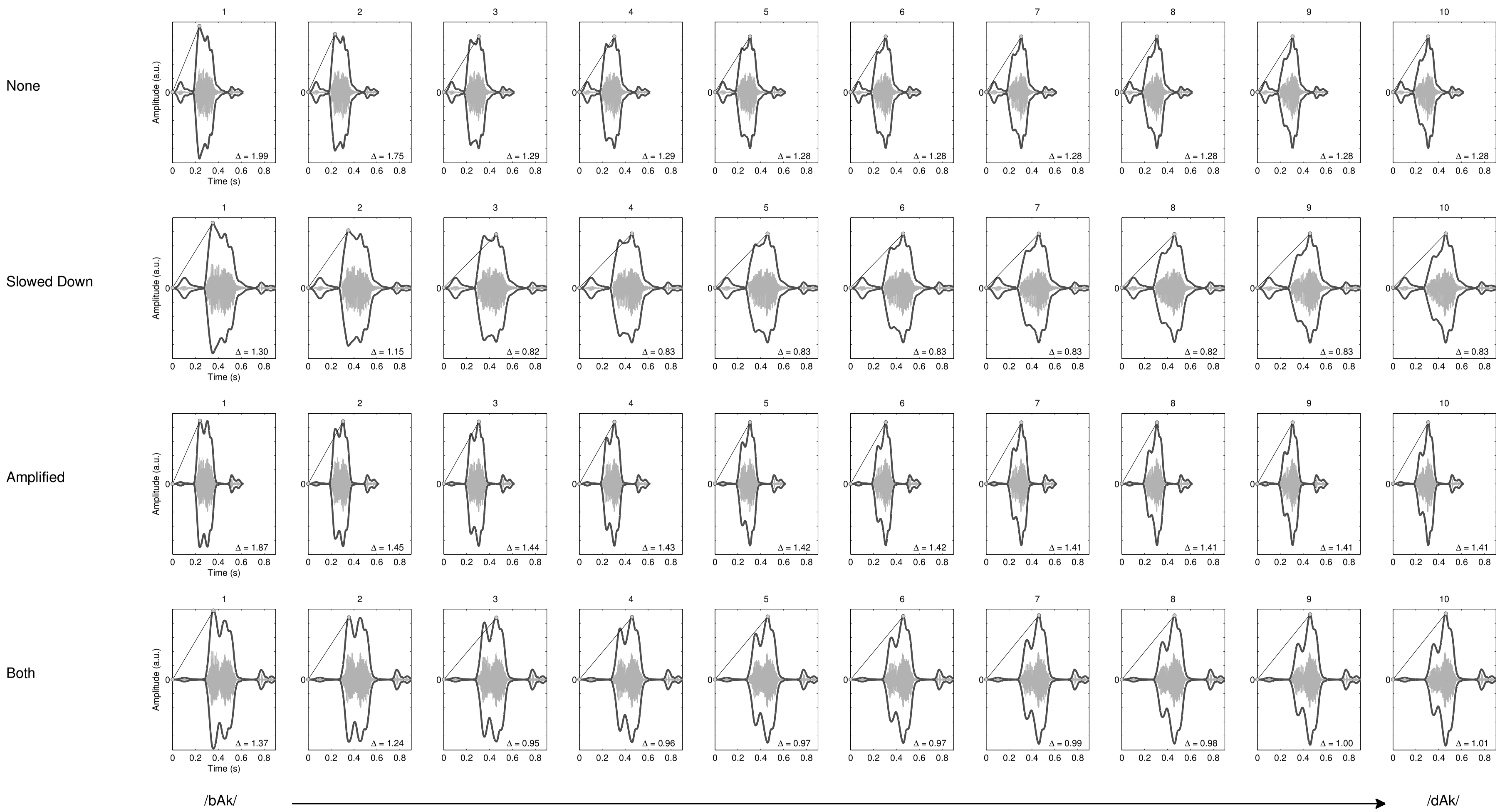


Figure 4.3 – Figures represent the smoothed envelope (exaggerated for clarity of presentation) of the amplitude waveform (smaller ghost image) of the stimuli used in the experiment. There are four manipulations of the 10-step continuum from /bAk/ to /dAk/. Envelope rise times (∆) are calculated as the slope of the line connecting the start of the stimulus onset amplitude to the maximum amplitude.

values for each stimulus are shown in Figure **4.4**. The figure reveals RFTe takes on different values for different steps on the continuum, but also across different acoustic manipulations.

Inharmonicity, or HNR measures how much energy in the spectrum is outside of the ideal harmonic sequence. To calculate this measure we assume the signal may be decomposed into a large number of partials, or sine waves that oscillate at a particular frequency. We also assume there is a fundamental frequency F0. The more harmonious the signal, the more it consists of partials that are multiples of F0. The Formants discussed earlier, can be considered such multiples. In an ideal situation, the second formant frequency F2 should be $2n * F0$, with $n = 2$. For the calculations presented here, the exact correspondence of the value of n to the order of the formant is not important as long as it is a multiple. Inharmonicity then represents how many of the partials in the signal are not multiples of F0, how much the signal deviates from an ideal harmonic sequence. This measure captures information about the impact of the changes in formant frequency with respect to the other formant frequencies present in the signal and might be a more accurate index of spectral changes than the absolute change in one formant such as the F2 slope. Table **4.1** lists the inharmonicity values of the stimuli as the percentage energy in non-harmonic partials. Again, there are clear differences between stimuli on the continuum and between the acoustic manipulations. The stimuli used were synthesised ((but based on actual recordings of utterances, see van van Beinum et al., 2005) to create a continuum in which the F2 onset frequency is the only major spectral change. F2 is constant at 1100 Hz in /bAk/ but the onset increases in ten steps to 1800 Hz in /dAk/. In the table it can be seen that /bAk/ is more inharmonious than /dAk/, which might seem counter-intuitive since in /bAk/ there is no change in F2 onset. However the fundamental frequency F0 of most of the stimuli is about 220 Hz, which yields about 1800 Hz with $n = 3$. The closest harmonic partial to 1100 Hz is 880 with $n = 2$.

Table **4.1**
Inharmonicity of the 40 Stimuli Used in the Experiment. The Numbers Represent Percentage of Energy in the Signal that is Outside of the Harmonic Sequence.

	Acoustic Manipulation			
Stimulus	*None*	*Slowed Down*	*Amplified*	*Both*
/bAk/	39.87	42.52	43.81	48.79
2	38.99	42.50	42.80	48.57
3	38.40	41.09	42.38	46.26
4	37.46	41.01	42.32	47.30
5	37.09	40.58	42.24	46.45
6	36.95	40.37	42.14	46.05
7	36.81	40.13	42.00	45.13
8	36.85	39.67	40.70	44.14
9	36.71	39.57	40.59	43.84
/dAk/	36.20	38.89	40.91	43.35

4.5 Complex Dynamic Pattern Measures: A Complexity Matching Hypothesis

A spectrogram representation of a speech sound (see Figure **4.2**) reveals the complexity of the speech signal by displaying how much the energy at different frequency bands changes over time. The spectrograms presented in Figure **4.2** are less noisy than recordings of actual speech produced by a human, they are partially synthetic. When trying to understand how humans perceive such

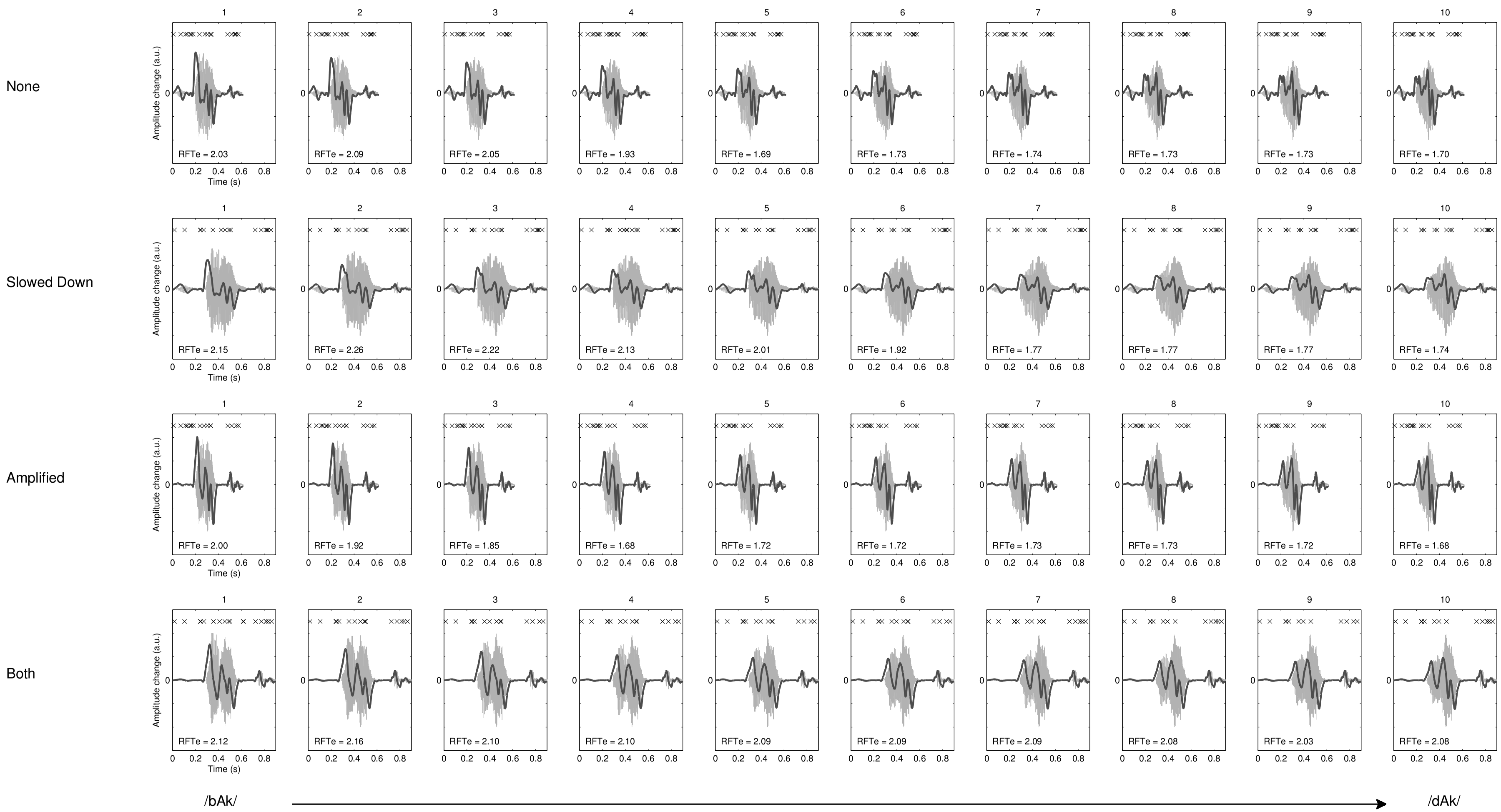


Figure 4.4 – Figures represent the derivative of the smoothed envelope (exaggerated for clarity of presentation) of the amplitude waveform (smaller ghost image). There are four manipulations of the 10-step continuum from /bAk/ to /dAk/. The derivative represents the rate of change of the envelope. The X markers in the top represent the zero-crossings of the derivative. The time between consecutive markers represents a rise or fall time. See text for details on calculating RFTe values.

a signal as a meaningful word or sentence it is tempting to focus on mechanisms that analyse frequencies or amplitudes and loose sight of the fact that the energy distribution in the spectrogram is a representation of a complex gesture, a motor action. In fact, there are at least 70 muscles involved in producing a simple syllable like /pa/ ranging from muscles that control respiration to the ones that control the tongue (Galantucci et al., 2006; Turvey, 2007). Producing speech sounds is very much a matter of sophisticated aerodynamic control by changing the shape of cavities air is forced to flow through (Porter and Hogue, 1998). The speech signal indeed appears to resemble the most complex dynamic motion known to physics, spatio-temporal chaos, or, turbulence: Models of human aspiration have been successfully validated against real turbulent airflow induced sounds generated in acoustic duct experiments (cf. Little and McSharry, 2007).

Perhaps the words of Horace Lamb, the author of the 1910 book: *The Dynamic Theory of Sound*, which is still in print today as an exact copy of the 1925 2nd edition (Lamb, 2004), should carry some weight. He was more famous for his work in hydrodynamics and is reported to have said:"*I am an old man now, and when I die and go to heaven there are two matters on which I hope for enlightenment. One is quantum electrodynamics, and the other is the turbulent motion of fluids. And about the former I am rather optimistic.*" (Moin and Kim, 1997). Indeed, the scientist who brought enlightenment on the subject of quantum electrodynamics, Richard Feynman, called turbulence:"*the most important unsolved problem of classical physics*" (cf. Moin and Kim, 1997). Lamb's dynamic theory of sound makes clear that a speech signal cannot be regarded as the vibration of a violin string propagating harmonic waves through the air (see Table **4.1**). A substantial part of the signal cannot be described as a harmonic sequence. The sound wave produced by a string is to the sound wave produced by a human speaker as a gentle summer breeze is to a hurricane.

Several authors have suggested aggregate, or collective levels of control that enable coordination of tasks with mind-boggling numbers of degrees of freedom such as speech perception and production. The uncontrolled manifold (Scholz and Schöner, 1999) and synergies (Turvey, 2007) are examples of such higher order mechanisms of control. They represent theoretical constructs based on a causal ontology in which interactions between components do the explanatory work for the theory, not the components themselves. The most sophisticated theoretical frameworks treat action and perception as a coupling of levels in a single complex system whose behaviour can only be explained as an inseparable whole (e.g., Chemero, 2009; Chemero and Turvey, 2007; Gibson, 1979; Michaels and Carello, 1981; Schoner and Kelso, 1988). Evidence is accumulating that humans are able to coordinate their behaviour by exploiting specific invariant properties of complex dynamical patterns either due to 'attraction to criticality' or 'complexity matching'. Attraction to criticality refers to the ubiquitous observation of 1/f scaling (pink noise) in time-series of human physiology and performance, which is associated with health and well being (cf. Goldberger et al., 2002), proficiency and fluency of performance (for example in motor learning (Wijnants et al., 2009), or as nested constraints on performance (Wijnants et al., 2012a)). Complexity matching is a remarkable synchronisation and coordination phenomenon in which participants are able match the complex scaling properties of an external stimulus in a record of their responses (e.g., finger tapping to a 'fractal' metronome Coey et al., 2014).

Formally, the terms fractal, power-law and scaling refer to different, related properties of mathematical objects, but in general fractal dynamics, power-law or 1/f scaling all refer to the observation of self-affine structure in empirical time series (cf. Van Orden et al., 2003; Kantelhardt, 2011). As shown in equation 4.3, self-affinity is different from self-similarity in that the similarity between small and large scale structures in time-series can only be observed by asymmetric scaling of the time axis t and value axis $x(t)$ by a factor a^H (cf. Kantelhardt, 2011). The scaling exponent H (or Hurst exponent) indicates factor that allows the self-affine structure to be observed as self-similar structure:

$$x(t) \to a^H x(at) \tag{4.3}$$

The scaling exponent can be associated to the fractal dimension of the signal or its generating process (see Hasselman, 2013, for a discussion of different scaling exponents and how they are related to fractal dimension). It is the invariant structure that is hypothesised to be exploited as a 'global' control variable, as if it were a complex resonance frequency (Aquino et al., 2011, ; Hasselman (2015)). Evidence of selective matching of dynamical behaviour to scaling exponents in different observables measured simultaneously throughout the body, suggests a complex multi-scale coupling relationship between physiological and psychological processes may exist (Rigoli et al., 2014). Complexity matching has also been reported for dyadic interactions, for example interpersonal coordination of coupled movements (Marmelat and Delignières, 2012) and overt behaviour during joint problem solving (Abney et al., 2014).

The important question for the present context of speech perception in average and dyslexic readers is whether the speech signal can be considered to reveal the invariant patterns and temporal complexity of which it is hypothesised listeners could exploit. The methods used in the studies that evidenced complexity matching as a phenomenon of perception, action and behaviour coordination were (Cross-) Recurrence Quantification Analysis (see e.g., Coey et al., 2014; Abney et al., 2014) and fractal analyses such as Detrended Fluctuation Analysis (see e.g., Marmelat and Delignières, 2012; Rigoli et al., 2014). RQA measures as well as the Hurst exponent have been applied to analyse naturally produced speech with the goal of detecting abnormal speech due to pathology or disease (Little and McSharry, 2007). These measures were successful in distinguishing between pathological and healthy origins of the recorded signal and were hypothesised to represent information at the level of non-linear and turbulent airflows generated by complex gestures of the human speech apparatus. Naturally produced speech sounds have also been shown to reveal 'attraction to criticality' at different levels of analysis and across many repeated productions of the same sound (Kello et al., 2008). As indicated in the introduction, studies have shown that a characterisation of response latencies is associated to dyslexic reading, Moreover, multifractal spectrum of reading times in connected text reading has been found to distinguish between reading fluency and proficiency in literate adults (see, e.g., Wallot et al., 2013). Based on these studies a Complexity Matching Hypothesis (CMH) can be formulated with regard to speech perception and reading ability.

Given few studies on scaling and fluency in developmental dyslexia have been conducted so far (Wijnants et al., 2012a; Holden et al., 2014), it would be premature to attempt to formulate a 'complexity matching' aetiology sufficient for explaining the many empirical phenomena associated with developmental dyslexia. Moreover, in the present study, the objects of complex signal analyses are not trial series of response latencies generated by participants, but the stimuli used in the experiment. Another difference is the difference in constraint on available response options, a binary choice versus pronunciation of a word. A modest conjecture would be to adopt the 'proportional continuum' assumption and suggest that any differences between dyslexic and average readers in labelling the stimuli should be the result of less stable, more variable continuous processes that lead up to the choice for one of the two options. From that perspective one would assume differences on labelling to be small, rather than large, but that is a common expectation of many competing claims (cf. Serniclaes and Sprenger-Charolles, 2003). One specification

The CMH states that listeners will use the dynamically invariant, self-affine structure of the speech signal to categorise and label speech sounds.

The relative novelty of employing these techniques to study the role of speech perception in proficient and impaired reading warrants a more elaborate explanation and discussion of the analyses used in this study.

4.5.1 Phase Space Reconstruction and Recurrence Quantification Analysis

Turbulence can be observed in any propagating medium and may be (partially) described as spatiotemporal chaos, or deterministic randomness in time and space simultaneously. As a consequence it is very difficult to accurately measure, model, forecast, or control turbulence in a medium. Even so, applying so-called embedding theorems allows for a reconstruction of the dynamics based on a record of the complex behavior. A well known theorem is Takens' theorem (after Dutch mathematician Floris Takens, see Takens, 1981) and it states that the *m*-dimensional attractor of a dynamical system may be reconstructed from a measured time series of a single observable dimension of that system. Due to the fact that the behaviour of the system is governed by interactions on many different spatial and temporal scales (interaction dominant dynamics), information about the dynamics of the whole system must be present in the dynamics of its parts. By using *m* delayed copies of the observed time series as surrogate dimensions one can reconstruct the phase space of the system and analyse an approximation of the attractor dynamics of the entire system. Takens' theorem ensures that the reconstructed attractor is topologically equivalent to the original attractor when all of the *m* dimensions of the system would have been observed (see Marwan et al., 2007, for a detailed explanation).

After phase space reconstruction analyses usually focus on quantification of the dynamics of the reconstructed attractor. A method commonly used for this purpose is Recurrence Quantification Analysis, (RQA, Marwan et al., 2007; Webber Jr. et al., 2009; Zbilut et al., 1998; Webber Jr. et al., 2005). RQA is a non-linear time series analysis technique that can quantify complex temporal patterns by means of analysing trajectories through state space and noting when trajectory coordinates are in each other's vicinity, when they can be said to be a state that is *recurrent*. In Figure **4.5** the attractor of the first 1024 samples of the transition part of the amplitude time series of stimulus 10 (/dAk/) is reconstructed in three dimensions. The time series for surrogate dimension *m* is shifted by τ samples for each extra surrogate dimension *m*. The values for τ and m are chosen so that the reconstructed attractor will represent maximal information in the measured series, but its exact value is in principle not relevant (mutual information is used to choose τ and a false nearest neighbour analysis to choose *m*, see Riley and Van Orden, 2005, for details). The coordinates in reconstructed state space in Figure **4.5** are not randomly jumping from one region to another, but trace periodic orbits through specific locations in the state space. When two coordinates fall within a radius ϵ the two coordinates are said to be recurrent. Sequences of multiple coordinates that are recurrent signify a trajectory in phase space that is being revisited by the system. It is a trajectory or a location in the state space the system is attracted to and these recurrent coordinates and the structures they form are the objects of analysis in RQA.

In Figure **4.5** trajectories are clearly visible as orbits around the denser centre of the state space. It is also apparent that the choice for a radius size will greatly influence which coordinates will be recurrent (see Schinkel et al., 2008). In general the radius, or threshold used in RQA is set to a number that yields 1-5% recurring coordinates (out of all theoretically possible recurring points given the size of the state space). The recurrent coordinates are recorded in a recurrence matrix visualised by a recurrence plot of which an example is shown in Figure 5. Since we are looking at recurrent trajectories of one system the time series of m-dimensional coordinates is evaluated against itself (auto-recurrence). For each coordinate pair a distance can be established and if that distance is smaller than the radius a black dot is plotted. The dot represents the fact that at some point in time the coordinate under consideration will be revisited by the system, approximately that is. This yields a recurrence plot that can contain horizontal and vertical line structures as well as individual recurrent points. Diagonal line structures represent a sequence of different coordinates (a trajectory through state space) that is revisited by the system, the proportion of recurrent points that form a diagonal line is quantified as determinism (DET). A vertical line structure signifies that system dynamics are attracted to a specific location in state space where it remains for a longer period of time. The proportion recurrent points that form a vertical line is called laminarity (LAM)

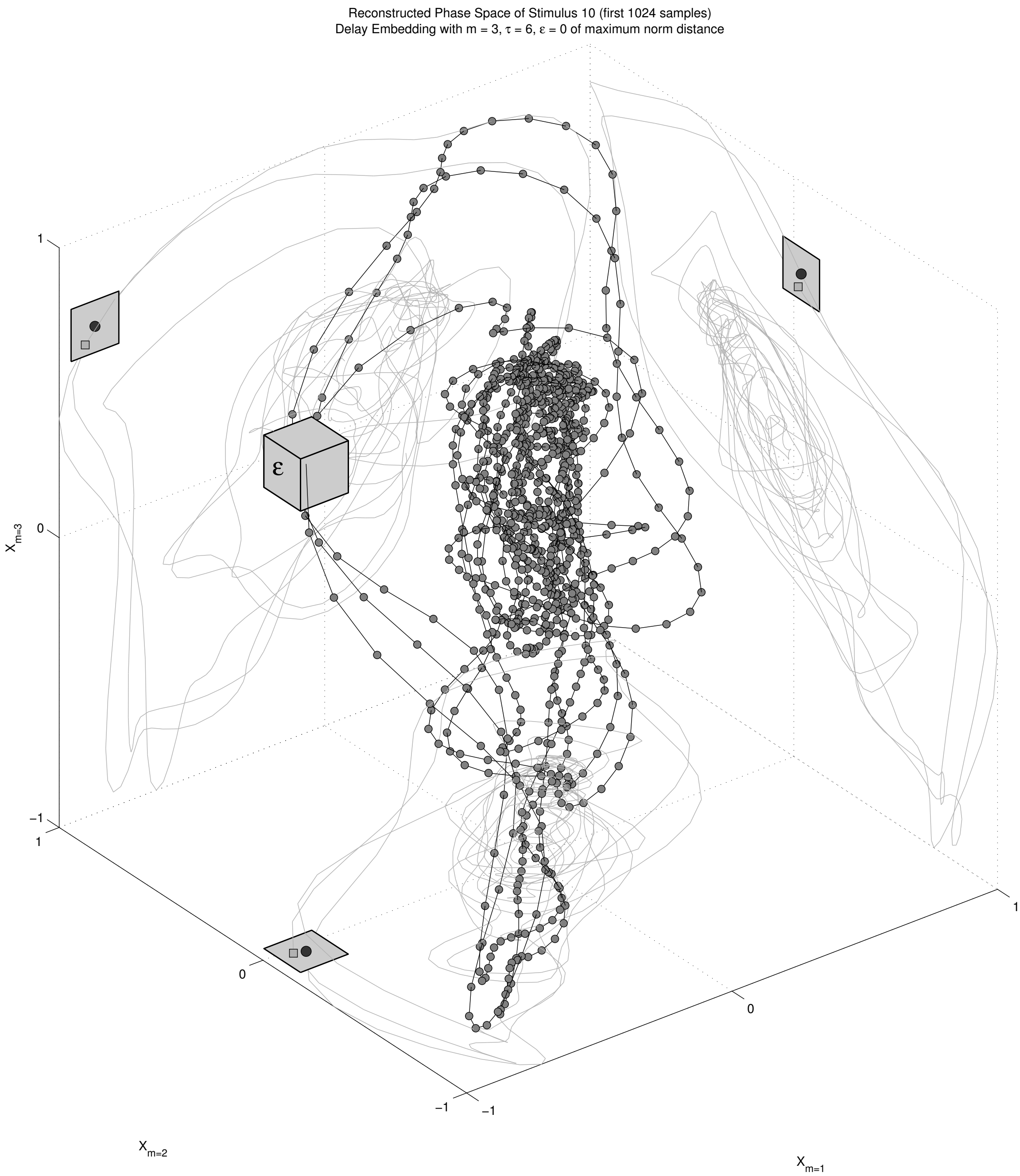
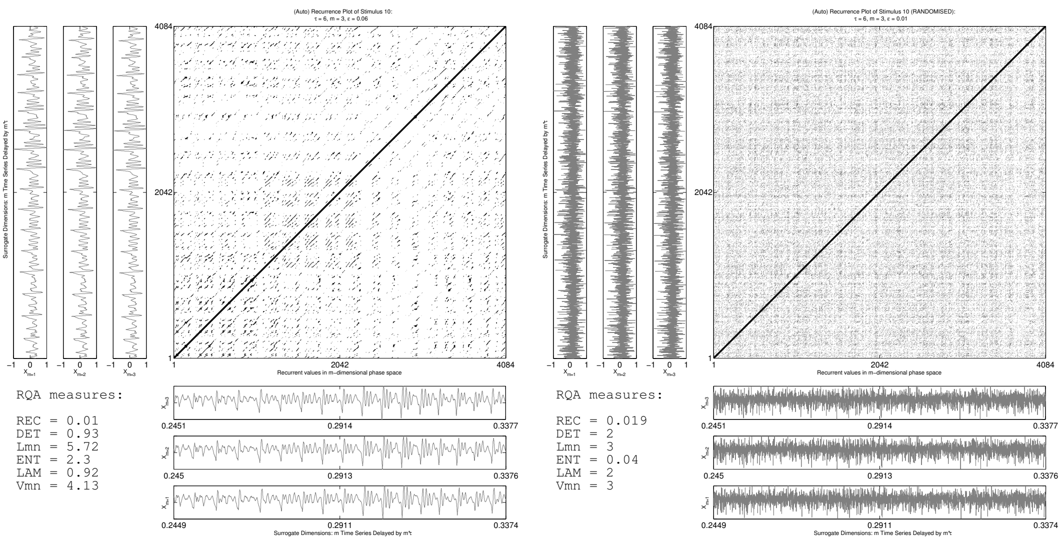


Figure 4.5 – A reconstruction of the 3D phase space of stimulus 10 (first 1096 samples) by the method of delay embedding. The planes show 2D projections of the time course of the surrogate dimensions created with an embedding delay $\tau = 6$. Points that fall within a distance ϵ (represented by the grey box for presentation purposes) will be plotted as recurrent points in the recurrence plot.



(a) Observed **(b)** Randomised

Figure 4.6 – A recurrence plot of the transition part of stimulus 10 (**4.6**a) and a randomly shuffled version (**4.6**b). Next to the recurrence plot axes are the surrogate dimensions *m* that span the phase space in which recurrent points are evaluated. They are offset by just $(m-1)*6$ samples.

and the mean vertical line length is called trapping time. One could say it quantifies whether the dynamics get 'trapped' in some region of the state space for a while. The plot is symmetrical about its diagonal, which represents the line of identity, or line of temporal incidence. By definition this is the longest line structure in the plot and is excluded from calculations.

The different line structures are clearly visible in the left pane of Figure **4.6** that is the recurrence plot of the entire reconstructed phase space of the transition part of stimulus 10, resampled to a length of 4096 datapoints of which the first 1096 are shown as a 3D reconstruction in Figure **4.5**. Figure **4.6**b is a randomised version of stimulus 10, the temporal order of the samples was randomised, destroying all the correlations that are in the data but retaining the same distributional properties (mean, variance, etc.). From the recurrence measures it can be seen that the recurrence rate in both panes (the number of recurrent points) is exactly the same. However, the measures that are calculated from the line structures that quantify the higher order recurrent patterns are very different. In the randomised plot all the DET and LAM disappeared, the temporal structure was destroyed even though the central tendency measures are exactly the same. This is a very basic test of whether the line structures are just accidental temporal alignments. A more sophisticated test would be to create spectral surrogates of the speech stimuli, or to do a bootstrap resampling on all the recurrence measures in order to create a confidence interval (cf. Schinkel et al., 2009). Figure **4.7** shows the recurrence plots for all the stimuli used in the present study. The threshold was varied in order to keep the recurrence rate exactly the same (at 10%) for all stimuli under consideration. Since we are looking at recurrences in reconstructed phase space, the assumption is that the figures represent the dynamical behaviour of the complex system that produced the speech signal.

RQA is used in an increasing number of studies across the different sub-disciplines the social and life sciences, such as motor development in infants (Aßmann et al., 2007), parent-child interaction (de Graag et al., 2012; Lichtwarck-Aschoff et al., 2012), syntactic coordination between child and caregiver (Dale and Spivey, 2006), dynamics of motor control (Wijnants et al., 2009, 2012a), cognitive constraints on postural stability (Shockley et al., 2003, 2007), eye-movements during conversation (Richardson et al., 2007), insight in problem solving (Stephen et al., 2009), and as a novel analysis tool in cognitive neuroscience (Bianciardi et al., 2007; Schinkel et al., 2007, 2009).

These quantifications are hypothesized to provide the best characterisation of the individual stim-

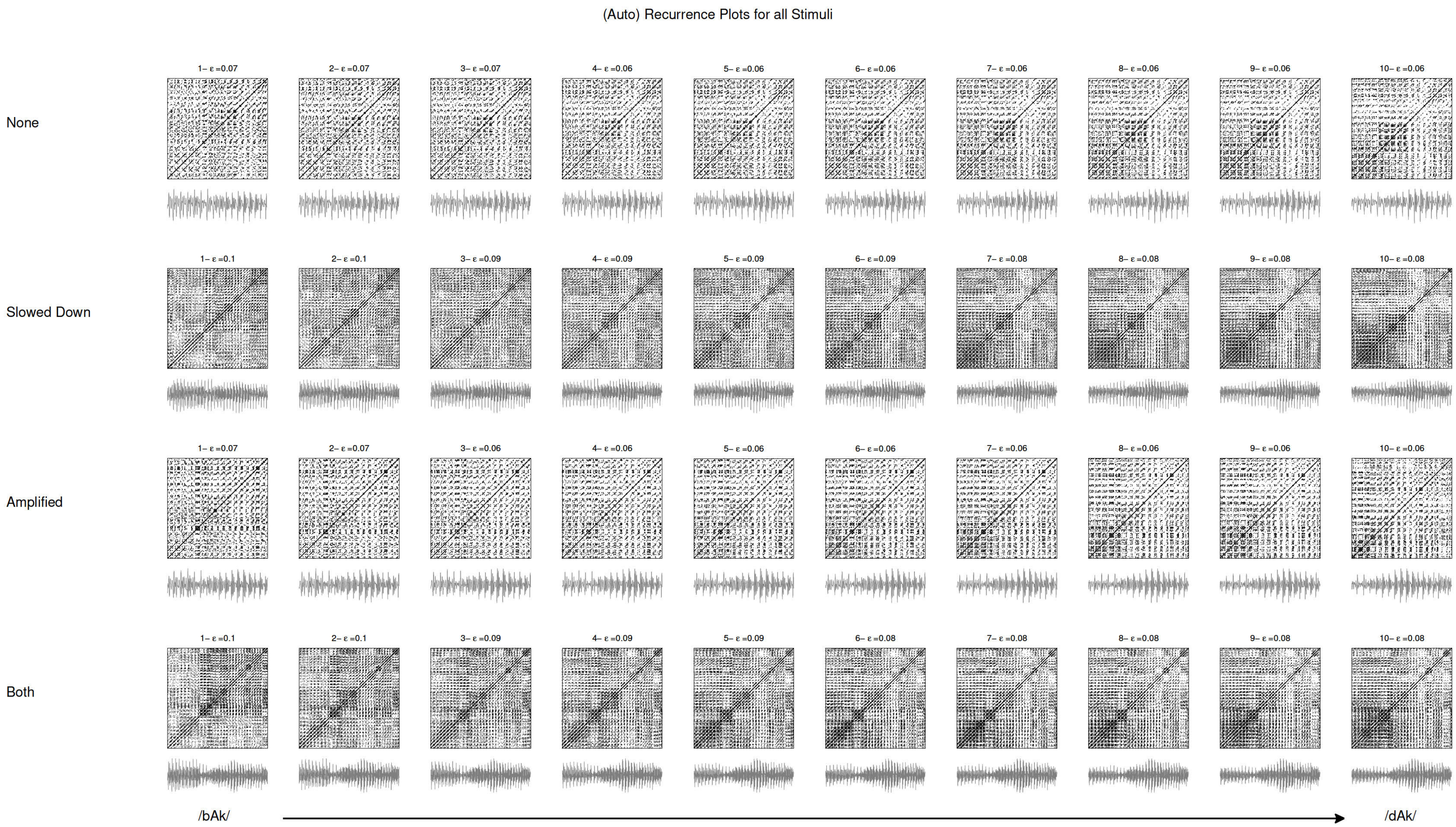


Figure 4.7 – Figures represent the recurrence plots of the amplitude waveform of the transition part of the stimulus (grey image below the plot). There are four manipulations of the 10-step continuum from /bAk/ to /dAk/. The plots were generated using embedding dimension (m) of 3 and an embedding delay (τ) of 6. The recurrence rate for each plot was kept constant by varying the radius (ϵ). This way the recurrence measures extracted from the plot are comparable across stimuli. See text for details.

uli that were artificially constructed to constitute an acoustic dimension and are therefore perceived to be mostly very similar. The relevant differences between the transition parts of the stimuli are expected to concern relative differences in patterns of sustained values (/bAk/) versus patterns of changing values (/dAk/). This is exactly what the non-redundant structures quantified by LAM and DET represent. Other measures calculated by RQA are mostly averages or maxima of the diagonal and vertical line structures (e.g., maximum diagonal line length or average diagonal or vertical line lengths) and were considered too homogeneous to characterise the individual stimuli.

Values used for reconstruction were $m = 3$ and $\tau = 6$ and the recurrence rate was kept constant at 0.1 (10%) by varying the radius ϵ (radius values are shown in Figure **4.7**; DET and LAM values are shown in Table **4.2**). As explained above, DET quantifies recurring trajectories through phase space and a high DET signifies a system that behaves very periodic and predictable. LAM quantifies recurrences of the system displaying the same type of behaviour, visiting the same region in phase space and staying there for a while. Some portion of the recurrent points quantified by DET will be representing laminar behaviour, so using a combination of these two measures in a classification analysis yields a description of the stimulus in terms of whether the dynamics are characterised by changing temporal patterns or patterns that stay relatively constant for some time.

Table **4.2**
Determinism and Laminarity of the 40 Stimuli Used in the Experiment. The Numbers Represent Proportion of Recurrent Points That Lie on Diagonal Lines (DET) or on Vertical Lines (LAM).

	Acoustic Manipulation							
	None		*Slowed Down*		*Amplified*		*Both*	
Stimulus	DET	LAM	DET	LAM	DET	LAM	DET	LAM
/bAk/	0.95	0.91	0.90	0.83	0.95	0.90	0.86	0.78
2	0.95	0.91	0.89	0.82	0.95	0.91	0.86	0.78
3	0.95	0.92	0.89	0.83	0.95	0.91	0.85	0.78
4	0.94	0.91	0.88	0.82	0.94	0.91	0.84	0.77
5	0.94	0.92	0.87	0.81	0.94	0.91	0.83	0.77
6	0.94	0.91	0.86	0.81	0.94	0.91	0.82	0.77
7	0.94	0.92	0.85	0.81	0.94	0.91	0.81	0.78
8	0.94	0.92	0.84	0.81	0.93	0.91	0.80	0.78
9	0.93	0.92	0.83	0.81	0.93	0.91	0.79	0.77
/dAk/	0.93	0.92	0.82	0.81	0.94	0.91	0.77	0.76

4.5.2 The Multifractal Spectrum

Fractal analyses are so-called variability analyses (cf. Bravi et al., 2011) that assess a scaling of 'bulk' with 'size' (Theiler, 1990). Expressed in terms of time-series it concerns the 'amount of fluctuation in a signal' $\approx$ 'scale at which fluctuation is quantified'. Figure **4.8** displays the steps in Detrended Fluctuation Analysis (DFA) in which the Hurst exponent is estimated by assessing a scaling of residual fluctuation (Root Mean Square variation) with bin size after detrending the binned signal. The top row of Figure 7 shows the envelope of the signal (black) and its 'profile' (grey). The profile is the cumulative sum of the signal after the mean has been subtracted. The following steps are applied to the profile (numbers refer to Figure **4.8**):

1. Divide the profile of length N into N_s non-overlapping segments v of size s (scale).
2. For each segment v of size s: Remove linear (or higher order) trend and calculate the RMS variation (residual variance).

3. The RMS variation of the variances calculated in step 2 represents a value of the fluctuation function $F^2(s,v)$ for the scale of size s.
4. Repeat 1-3 for increasing values of s (in the present study the minimum scale was 64, the maximum 4096). The slope of the fluctuation function $F^2(s,v)$ is the global scaling exponent H

In many empirical time series the scaling behaviour is multifractal rather than monofractal, that is, the signal is better characterised by a spectrum of local scaling exponents than a single global exponent (cf. Kantelhardt, 2011). Multifractal Detrended Fluctuation Analysis (Kantelhardt et al., 2002) is a generalisation of DFA that quantifies different orders of fluctuation, the q-order fluctuation of generalized moments. Standard DFA calculates the fluctuation function of the variance σ^2, which the second-order moment of a distribution of values (q=2). The standard deviation σ^1 is the first-order moment (q=1). Rewriting the familiar formulas for the standard deviation (root mean square deviation) and the variance (mean squared deviation) of a sample of observations, their relation to q-order fluctuation analysis is as follows:

$$\text{root mean squared deviation: } \sigma^1 = \sqrt[2]{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2} \;\rightarrow\; F^1(s,v) = \left[\frac{1}{N_s}\sum_{v=1}^{N_s}\sigma_v^2\right]^{\frac{1}{2}} \tag{4.4}$$

$$\text{mean squared deviation: } \sigma^2 = \sqrt[1]{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2} \;\rightarrow\; F^2(s,v) = \left[\frac{1}{N_s}\sum_{v=1}^{N_s}\sigma_v^2\right]^{\frac{2}{2}} \tag{4.5}$$

$$\text{q-order deviation: } \sigma^q = \sqrt[q]{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2} \;\rightarrow\; F^q(s,v) = \left[F^2(s,v)\right]^{\frac{q}{2}} \tag{4.6}$$

$$\text{q-order fluctuation: } F_q(s) = \left\{\frac{1}{N_s}\sum_{v=1}^{N_s}\left[F^2(s,v)\right]^{\frac{q}{2}}\right\}^{\frac{1}{q}} \tag{4.7}$$

Using q=2 in equation 4.7, will yield the RMS deviation of the variance. The q-order takes on the role of a zoom-lens for fluctuations: By increasing q to more positive values, large residual variances will be given more weight than smaller ones when establishing the scale dependency of the fluctuations in the signal. On the other hand, decreasing q to lower negative values has the opposite effect and will zoom in on the scale dependency of small residual variances.

To obtain a spectrum of scaling exponents for each q-order, the 4 steps of standard DFA are repeated for a q-continuum, which typically ranges from -10 to 10. The left column of Figure **4.9** shows for the 4 x 10 stimuli their fluctuation functions of order q = [-5, -2, 0, 2, 5]. The black dotted power law at q = 2 represent the fluctuation function of stimulus 1 that is show in the bottom row of Figure **4.9**. For each of the 40 stimuli, a 101 step q-continuum was estimated ranging from q=-10 to q=10 (including q=0). The scaling exponents H(q) are the slopes of those 101 fluctuation functions (Table S1 lists for each stimulus the average and SD of the norm of the residual after regression). Those slopes are plotted against q in the middle column of Figure **4.9**. If the stimuli were monofractals, there would have been no dependence of the scaling exponent $H(q)$ on the q-order for which it was calculated. The plots in the middle column of Figure **4.9** would all have been horizontal lines (see e.g., Figure 1d in Kantelhardt et al., 2002, p. 94). Here, it is clearly the case that all the stimuli used in the study should be considered multifractal signals. The multifractal spectrum (right column of Figure **4.9**) is a representation of the generalized scaling exponents (now called

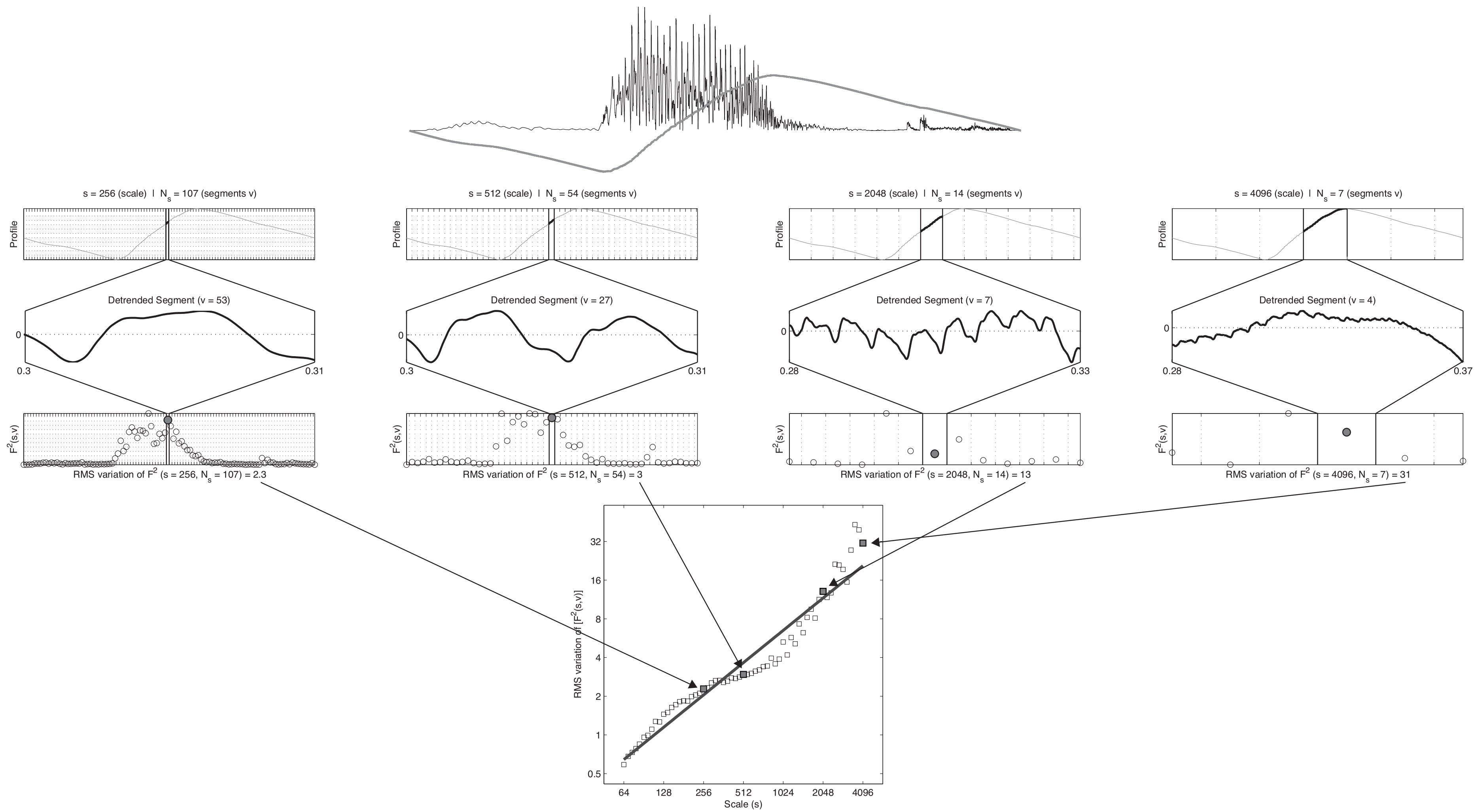


Figure 4.8 – Steps in Detrended Fluctuation Analysis. The profile of stimulus 1 is divided into nonoverlapping segments v of size s; the signal is detrended and the mean RMS deviation of the RMS variation in each segment is calculated; this is repeated for different scales and results in the fluctuation function: $F^2(s, v)$. The slope of the best fitting line through these points is an estimate of the Hurst exponent. See text for details.

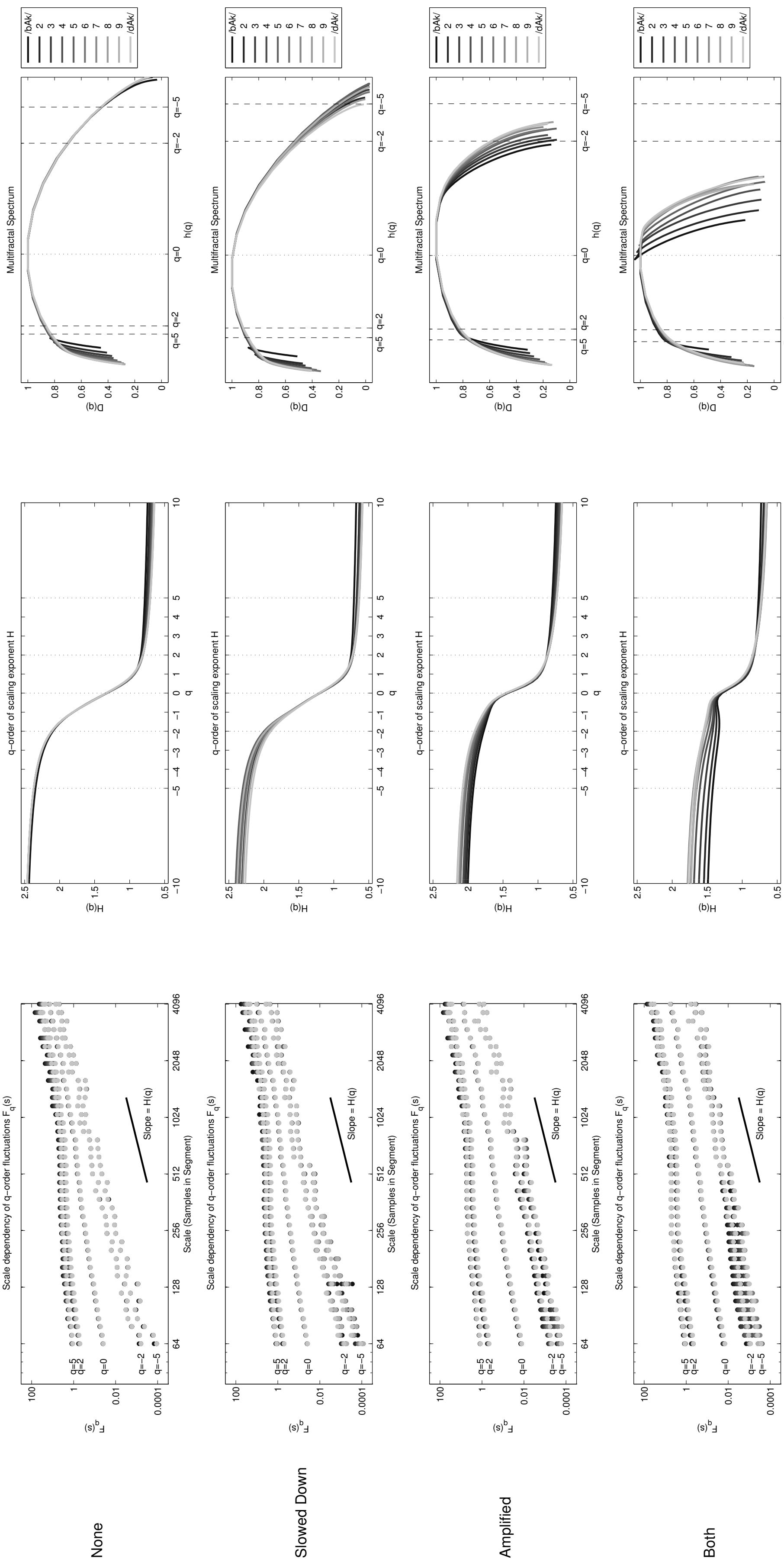


Figure 4.9 – Multifractal Detrended Fluctuation Analysis of the 40 stimuli.

singularity, Hölder, or generalized Hurst exponents) against $D(q)$, the q-order singularity dimension (the calculation of $D(q)$ is not shown here, see Ihlen, 2012, for details).

The multifractal spectrum does not need to be symmetrical and Figure **4.9** reveals that the discrepancy between stimuli may be revealed by considering the dispersion of h(q) separately for q-orders < 0 and q-orders > 0. As noted by Kuznetsov and Wallot (2011), each half of the singularity spectrum conveys different information about scaling properties of the signal. The measures of interest will be for the Coefficient of Variation for each half-spectrum:

$$\text{CVhq+} = \frac{s_{h(q>0)}}{\bar{h}(q>0)} \tag{4.8}$$

$$\text{CVhq-} = \frac{s_{h(q<0)}}{\bar{h}(q<0)} \tag{4.9}$$

For $q < 0$ (CVhq-) and $q > 0$ (CVhq+). Table **4.3** shows the values of the multifractal CV for each stimulus.

Table **4.3**
Coefficient of Variation of Local Scaling Exponents Calculated for $q < 0$ (Zooming in on the Scale Dependency of Smaller Residual Variation) and $q > 0$ (Zooming in on Scale Dependency of Larger Residual Variation)

	Acoustic Manipulation							
	None		*Slowed Down*		*Amplified*		*Both*	
Stimulus	CVhq-	CVhq+	CVhq-	CVhq+	CVhq-	CVhq+	CVhq-	CVhq+
bAk/	0.118	0.308	0.112	0.385	0.135	0.170	0.116	0.124
2	0.131	0.330	0.135	0.387	0.149	0.180	0.144	0.142
3	0.142	0.341	0.146	0.421	0.165	0.184	0.161	0.162
4	0.152	0.348	0.156	0.441	0.179	0.192	0.176	0.192
5	0.162	0.354	0.164	0.452	0.190	0.211	0.191	0.204
6	0.170	0.359	0.170	0.434	0.200	0.208	0.194	0.204
7	0.176	0.363	0.172	0.449	0.205	0.212	0.195	0.199
8	0.182	0.365	0.174	0.417	0.208	0.226	0.195	0.161
9	0.185	0.366	0.176	0.449	0.210	0.220	0.189	0.166
/dAk/	0.185	0.366	0.176	0.399	0.210	0.218	0.188	0.182

4.5.3 Which measure do participants use to identify /bAk/ and /dAk/?

A recent successful application of RQA and other complexity measures to speech sound classification was done in the context of voice disorder detection (Little and McSharry, 2007). Natural recordings from a database of more or less clear examples of voice disorders were analysed on the classification ability of several measures thought to be theoretically important to detect the voice disorders (jitter, shimmer, amplitude irregularity, and HNR). These classical measures, together with the complexity measures Recurrence Period Density Entropy (RDPE, a measure derived from the recurrence times in the plot) and a normalised scaling exponent (Hnorm, derived from Detrended Fluctuation Analysis; DFA) were evaluated for their classification performance in a quadratic discriminant analysis (QDA). The complexity measures were superior in distinguishing between normal and voice disorder recordings (overall classification 91.8% correct for RDPE/Hnorm with other measure pairs ranging from 76.4% to 81.4%; see Table 1 in Little and McSharry, 2007).

In this study I will use a similar approach to categorise the speech signals as Little et al. (2007) did, but the targets for the quadratic discriminant analysis (QDA) will not be disordered speech vs.

healthy speech, but the observed labelling of the stimuli by average and dyslexic readers as either /bAk/ or /dAk/. The labelling patterns will be experimentally assessed by administering a labelling task of four versions of a 10 step /bAk/ to /dAk/ continuum (None, Slowed Down, Amplified and Both). A first research question is whether there are differences in labelling between experimental groups and stimulus types. This could potentially yield eight different labelling patterns. If there is a difference between experimental groups, QDA will be performed for each group separately. The features used by QDA to classify the stimuli will be the measures discussed above. These measures are extracted from one and the same set of stimuli, but represent different theoretical perspectives on (impaired) speech perception. Figure **??** and Table **4.4** summarise the different hypotheses (AT-PDH, RTPDH, CMH) and associated measures extracted from different representations and quantifications of the temporal patterns in the speech signal. The simple main hypothesis is that the combination of measures that yield the best classification performance is the most likely source of information used by the participants in this study to label the stimuli.

Table **4.4**
A Summary of Hypotheses about Important Elements of the Speech Signal for Speech Perception and Associated Measures.

	Signal		**Measure**		**QDA**
Hypothesis	*Representation*	*Transform*	*Name*	*Type*	*Acronym*
RTPDH	Time-frequency	Short-time Fourier	2^{nd} formant slope	Component Process	F2
		Short-time Fourier	Inharmonicity	Periodicity	NHR
ATPDH	Analytic signal	Hilbert transform	Slope to max. envelope	Component Process	maxENV
	Hilbert transform	Rise	fall time entropy	Periodicity	RFTe
CMH	State space	Delay embedding	Recurrent trajectory	Complex Pattern	LAM / DET
	Scale space	Multifractal spectrum	Multifractal CV	Complex Pattern	CVhq+ / CVhq-

4.6 Method

4.6.1 Data Sharing and Reproducibility of Results

The raw and aggregated data, stimulus files and Matlab code (The MathWorks, 2012) to reproduce the analyses and figures in this article are available at the Open Science Framework: `https://osf.io/a8g32`. The files are annotated and demonstrate how to extract the stimulus features from the audio files, how to create figures and perform the QDA analysis. In addition, the raw data is available in spreadsheet format.

4.6.2 Participants

Children could enter the study as participants after their caregivers signed an informed consent form (equivalent to “Consent Form 4 - Under 12” issued by the Ethics Committee of the Faculty of Social Sciences of the Radboud University Nijmegen. An English translation is available in the supplementary materials). There were 80 participants (age range 101.2 to 159.3 months) from 9 different schools in the south-east of the Netherlands. Half of the subjects (40) were dyslexic readers as indicated by two reading tests: A timed-reading task for regular words “Drie-Minuten-Toets”; Verhoeven, 1995 and a timed pseudo-word reading task (“KLEPEL”; Van den Bos et al., 1994). When the child’s scores on both tests were within the 25th percentile (norm score by age), the child was considered to have severe reading problems. For one participant who completed the study, no data was recorded in the output file and could not be included. Table **4.5** displays the information for the participants whose data were analysed (all data are available in the supplementary materials).

Table **4.5**
Results For the Two Groups of Children Participating in the Experiment. The DMT Scores Represent Words Read Correctly in One Minute. Level Of Difficulty Increases From DMT1 to DMT3. KLEPEL Represent Correctly Read Pseudowords In Two Minutes.

	Average Readers		**Dyslexic Readers**	
	Mean	*SD*	*Mean*	*SD*
Age (months)	127.2	12.3	133.5	14.9
DMT1	100.0	15.5	72.1	15.7
DMT2	94.6	18.2	60.05	15.7
DMT3	84.3	16.7	48.0	16.2
KLEPEL	74.1	17.4	32.4	12.7
Gender	22 Boys	18 Girls	19 Boys	20 Girls
N	40		39	

4.6.3 Stimuli and Acoustic Manipulations

The stimuli were based upon natural speech recordings for the words /bAk/ [container] and /dAk/ [roof] and transformed to create a 10-step /bAk/ to /dAk/ continuum (van Beinum et al., 2005) using the Praat program (Boersma and Weenink, 2002). The stimuli differed only with respect to the second formant transition of which the onset frequency was gradually increased from /bAk/ to /dAk/ (see Table **4.6** for exact values). All the stimuli on this F2 continuum were manipulated in three manners using the Praat program (Boersma and Weenink, 2002). First, the speech signal was Slowed Down to 150% of its original length. This was achieved by a Pitch Synchronous Overlap and Add (PSOLA) algorithm (see e.g.. Segers and Verhoeven, 2005). Second, the signal was Amplified by 20 dB, for the fast changing spectral elements. The algorithm used to do this in Praat was similar to the one used by (Nagarajan et al., 1998), who confirmed this in a personal communication with Segers and Verhoeven (2005). Third, Both manipulations were applied as is done in the FastForWord program (Merzenich et al., 1996; Tallal et al., 1996): the speech signal was slowed to 150% of its original length and all the fast transitional elements were then amplified by 20 dB. There was of course also a continuum which had None of the manipulations applied to it. This yielded 40 different stimuli in total.

4.7 Procedure

4.7.1 Speech Perception Experiments

The speech identification task (labelling task) was presented on a laptop computer in a quiet room at the children's school. There were two tasks conducted in two sessions; an identification task (reported in this article) and a discrimination task (reported in Hasselman, 2014b). In the identification task, the participants were asked to rest their left and right index fingers on a coloured key on the left side [z] and right side [/] of the keyboard. After an attentional beep and fixation cross a smiley face appeared on the screen, which then uttered a word, one of the stimuli. The cover story was that the smiley face could not speak very well and the child had to help find out which out of two possible words (/bAk/ [roof] or /dAk/ [container]) it had just said. After the utterance of the word two frames appeared on the screen, one on the left, one on the right with either a picture of a roof or a container inside (positions were randomised). The child had to press the button corresponding to the position of the picture named by the smiley face. Prior to the experimental trials, 10 practice trials were presented using different pictures and pronunciations that were all clear exemplars. Feedback was given on the responses during these practice trials and no child made more than 3 errors during practice. During the experimental condition, the unmanipulated and the three types of manipulated /bAk/ and /dAk/ stimuli were presented in a random order. Each stimulus was presented twice resulting in 80 stimulus presentations (2 x 4 manipulations x 10 stimuli). The stimulus materials (audio files and pictures) are available in the on-line supplemental information.

4.7.2 Extracting the Stimulus Characteristics

The 40 stimuli were 16 bit digital audio files in .WAV format, with a sample rate of 44.1 KHz. These were always used as the basis for extracting the following measures:[1] The slope of the second formant transition (F2 slope, Figure **4.2**), the time it took for the envelope to reach its maximal value (mxENV Slope, Figure **4.3**), the entropy of rise and fall times (RFTe, Figure **4.4**). Settings were used in Matlab that mimic the default behaviour of the Praat program (Boersma and Weenink, 2002) so the output of this script should be similar to output generated by Praat. For the Inharmonicity measure (HNR; Table **4.1**) and the measures obtained from recurrence quantification analysis (Figure **4.7**) only the transition part of the stimulus was considered. Following Little and McSharry (2007), to assure that the RQA is performed on time series of equal length, all files were resampled to 4096 samples (waveforms shown under the RP plots in Figure **4.7**). The Multifractal spectrum was obtained by Multifractal Detrended Fluctuation Analysis based on the entire stimulus signal, using Matlab code by Ihlen (2012).

4.7.3 Statistical Analysis

For each participant there were 80 responses of either /bAk/ or /dAk/. These data were entered in a logistic multilevel model (using MLwiN version 2.2 Rabash et al., 2009) with the 80 measurement occasions representing responses to a random permutation of the ordered F2 continuum at level1. The responses at the level of the measurement occasions were considered binomially distributed as 0 and 1 and a logit link function was used. The repeated measurements can be thought of as clustered within the participant, who represent a second level of random variation in the model (level2). The modelling strategy was as follows: First it was examined whether the multilevel model gave a better fit than a single level model with just measurement occasion defined as a level. Then, the empty multilevel model for change was fitted (M_0), which in the present case means that a zero inflated fixed effect predictor was added representing the stimulus rank order on the continuum (0-9). In a subsequent model (M_2) it was examined whether stimulus rank could explain random variation in the slopes of the curve at the level of the participants (level2). If so, this means the variation in labelling of the continuum between participants can be understood as random variation with respect to the average labelling curve of the entire sample. In the next step (M_3) level1 and level2 covariates were added: A dummy variable that represents the four stimulus types (level1), and a dummy variable that represents whether subjects are dyslexic or average readers (level2). In the final modelling step (M_4) various interactions were tested including cross-level interactions between participant type and stimulus type. The models were fitted using a Monte Carlo Markov Chain simulation with 150,000 iterations (Browne et al., 2009). This number was chosen after inspecting the Raferty-Lewis diagnostic for each parameter estimate at each modelling step and was found to yield a very safe margin for all predicted parameters.

The predictions of the logistic multilevel model for each stimulus were used as targets for the quadratic discriminant analysis (QDA). If the lower 95% confidence bound predicted by the logistic multilevel model exceeded the chance level of 0.5 it was noted for that stimulus that /dAk/ was

[1] Extraction of these measures is described in detail in the supplementary materials. Many functions are based on freely available Matlab scripts, all of which are documented in the file *Hasselman2014-extractmeasures.m* available here: `https://osf.io/a8g32/files`

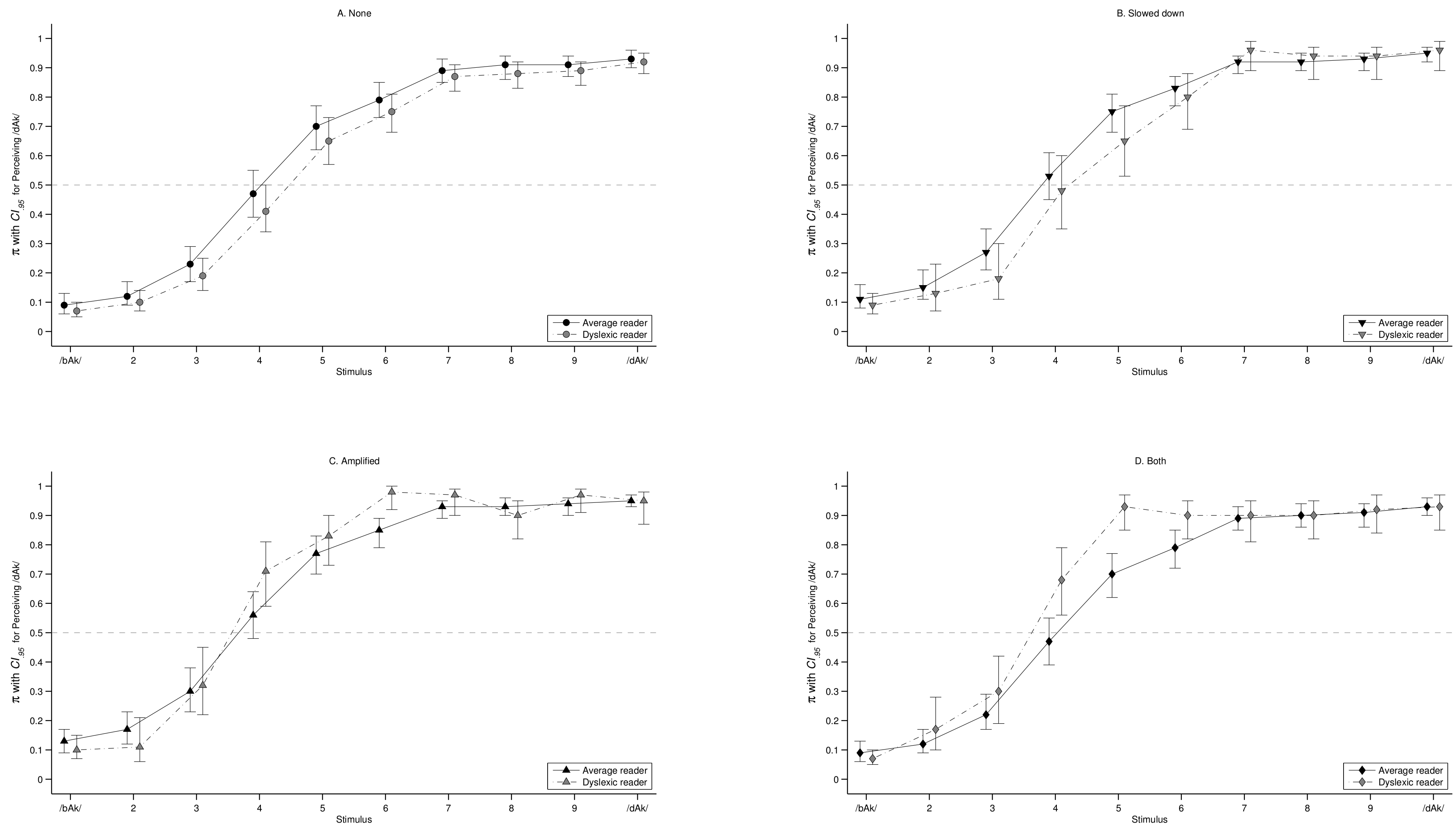


Figure 4.10 – Predicted probability of perceiving /dAk/ with 95% CI for each stimulus on the artificial continuum (predictions based on 150,000 MCMC replications of model M_4 in Table **4.5**). The lines summarise average and dyslexic readers and panels represent each type of manipulation: A. *None*; B. *Slowed Down*; C. *Amplified*; D. *Both*. The points are offset around the stimulus number on the x-axis to increase readability. There are two clear instances of non-overlapping confidence intervals (Panel C, stimulus 6; Panel D, stimulus 5). Values for the entire sample (Model M_3) are given in Table **4.7**.

Table **4.6**
Model Evaluation With Identification Label (idL) As Dependent Variable. The Bayesian Deviance Information Criterion Was Used For All Consecutive Models Estimated with MCMC (150,000 iterations). D = Posterior Mean Deviance, $D(\phi)$ = Deviance of Posterior Means, $pD(D - D(\phi))$ = Effective Number of Parameters, DIC = Deviance Information Criterion. See Text For An Explanation of the Modelling Steps.

	M$_{single}$		**M$_0$**		**M$_1$**		**M$_2$**		**M$_3$**		**M$_4$**	
idL$_{ij}$ =	β	$S.E.$	β	$S.E.$	β	$S.E.$	β	$S.E.$	β	$S.E.$	β	$S.E.$
Fixed Part												
Intercept	0.51	0.03	0.53	0.04	-2.04	0.10	-2.39	0.15	-2.77	0.19	-2.59	0.20
stimulus					0.66	0.02	0.77	0.04	0.8	0.05	0.78	0.04
Slowed Down (D1)									0.29	0.10	0.16	0.14
Amplified (D1)									0.63	0.10	0.37	0.14
Both (D1)									0.36	0.10	0.01	0.14
Dyslexic (D2)											-0.33	0.20
Slowed Down X Dyslexic											0.25	0.21
Amplified X Dyslexic											0.53	0.21
Both X Dyslexic											0.73	0.20
Random Part												
Level 2												
Intercept variance			0.11	0.03	0.36	0.08	1.79	0.38	1.81	0.39	1.83	0.38
Slope variance							0.42	0.09	0.42	0.10	0.42	0.09
Intercept-Slope covariance							0.12	0.03	0.12	0.03	0.12	0.03
Level 1												
Binomial variance	$var(idL_{ij} \mid \pi_{ij}) = \frac{\pi_{ij}(1-\pi_{ij})}{1}$											
D	8360.07		8220.42		5335.96		4958.18		4920.15		4907.22	
$D(\phi)$	8359.09		8167.94		5271.46		4835.14		4793.9		4777.51	
$pD(D - D(\phi))$	0.98		52.47		64.5		123.04		126.24		129.71	
DIC	8361.06		8272.89		5400.45		5081.22		5046.39		5036.93	

perceived. Otherwise the target for the discriminant analysis was /bAk/ for that stimulus. This resulted in a string of 40 zeroes and ones. The objective of the discriminant analysis was to replicate the classification in zeroes and ones based on pairs of the measures discussed above. The following pairs were tested mxENV Slope / F2 Slope; HNR / F2 Slope; RFTe / mxENV Slope; RFTe / HNR; LAM / DET. The pairs were all converted to the unit scale before analysis. The algorithm used to perform QDA was the same as described in Little and McSharry (2007). This procedure allows for calculation of 95% Confidence Intervals around the percentage correctly classified stimuli by bootstrap resampling. All QDA analyses were based on 15,000 bootstrap replications.

4.8 Results

4.8.1 Multilevel Logistic Model

The results of multilevel modelling taking the individual trials of the identification experiment as the dependent variable at level1 and subjects at level2 are shown in Table **4.7**. A graphical representation of the predictions by the final model is shown in Figure **4.10**. In the final model, there was no significant main effect of experimental group (dyslexic reader vs. average reader), but there was a significant cross-level interaction between experimental group and acoustic manipulation. This interaction is revealed in Figure **4.10** where in Panel C (Amplified) and D (Both) there two clear examples of non-overlapping CI between the labelling curves of average and dyslexic readers for

Table **4.7**

*Predicted Probability (π) for Perceiving /dAk/ for all Participants from MCMC Model Estimation (Median of 150,000 iterations yielding 95% CI) for each stimulus and Acoustic Manipulation (M_3 of Table **4.6**). When the Lower CI Limit Exceeded 0.5 the Target for QDA was /dAk/, Otherwise it was /bAk/.*

Formant Onset (Hz)				**Predicted Probability (π) for Perceiving /dAk/ per Acoustic Manipulation**							
Stimulus	F1	F2	F3	*None*	95% CI	*Slowed*	95% CI	*Amplified*	95% CI	*Both*	95% CI
/bAk/	440	1100	2700	0.06	(0.04, 0.08)	0.08	(0.05, 0.12)	0.13	(0.07, 0.24)	0.18	(0.08, 0.35)
2	|	1178	|	0.12	(0.08, 0.17)	0.15	(0.09, 0.25)	0.25	(0.13, 0.43)	0.32	$(0.15, 0.57)^{U}$
3	|	1255	|	0.23	(0.16, 0.33)	0.28	(0.17, 0.44)	0.43	$(0.24, 0.64)^{U}$	0.51	$(0.27, 0.76)^{\pi}$
4	|	1333	|	0.40	$(0.28, 0.54)^{U}$	0.47	$(0.29, 0.66)^{U}$	0.62	$(0.38, 0.81)^{\pi}$	0.70	(0.42, 0.88)
5	|	1411	|	0.60	$(0.44, 0.74)^{\pi}$	0.66	$(0.45, 0.82)^{\pi}$	0.78	$(0.56, 0.91)^{L}$	0.84	$(0.60, 0.95)^{L}$
6	|	1489	|	0.77	$(0.61, 0.87)^{L}$	0.81	$(0.63, 0.92)^{L}$	0.89	(0.72, 0.96)	0.92	(0.75, 0.98)
7	|	1567	|	0.88	(0.76, 0.94)	0.91	(0.77, 0.96)	0.95	(0.84, 0.98)	0.96	(0.86, 0.99)
8	|	1644	|	0.94	(0.86, 0.98)	0.95	(0.87, 0.98)	0.98	(0.91, 0.99)	0.98	(0.92, 1.00)
9	|	1722	|	0.97	(0.93, 0.99)	0.98	(0.93, 0.99)	0.99	(0.96, 1.00)	0.99	(0.96, 1.00)
/dAk/	440	1800	2700	0.99	(0.96, 1.00)	0.99	(0.97, 1.00)	0.99	(0.98, 1.00)	1.00	(0.98, 1.00)

^{L} Lower CI limit $\geq$ 0.5 threshold (used as observed classification boundary)
$^{\pi}$ Predicted Median Probability $\geq$ 0.5 threshold
^{U} Upper CI limit $\geq$ 0.5 threshold

stimulus 6 in Panel C and stimulus 5 in Panel D.

In both cases the dyslexic readers have a higher odds for perceiving /dAk/. Another difference between the groups may be observed when evaluating at which stimuli the lower confidence bound of the odds for perceiving /dAk/ exceeds the chance level of 0.5. Again the difference between the groups is observed with stimuli of category Amplified and Both (Panel C and D in Figure **4.10**). The dyslexic readers' odds for perceiving /dAk/ is with 95% certainty higher than chance at stimulus 4 for these manipulations, whereas for normal and Slowed Down manipulations it is at stimulus 5. For average readers this boundary is always at stimulus 5 irrespective of the acoustic manipulation. In Table **4.7** the significant parameter estimates of the final model (M_4) corroborate this: At each unit step increase in F2 frequency (stimulus number) there is an increase in the odds of perceiving /bAk/. Amplified stimuli also increase the odds of perceiving /dAk/ and for the group of dyslexic readers Amplified and Both stimulus types add even more to those odds. The random intercept and slope variance indicate that labelling curves vary across participants. Adding predictors and cross-level interactions did however not noticeably decrease, or explain this variance (changes are in 3rd decimal of estimated parameters). The DIC statistic did decrease with each consecutive model indicating a better model fit.

4.8.2 Quadratic Discriminant Analysis

Because the outcomes of the multilevel logistic model yield different boundaries at which dyslexic and average readers switch from /bAk/ to /dAk/ for stimuli of type Amplified and Both, the QDA was performed for each group separately using these labels as the target for the classification. At the same time, there was no significant main effect of group and the boundaries for the entire sample as predicted by M_3 (see Table **4.6**) deviated from the boundaries predicted by M_4 for each group. To investigate the impact of these differences an additional QDA classification was performed using the predicted labels on the level of the sample. The results for the sample are shown in Figure **4.11** and Table **4.7** that also includes the results for the predicted labels of M_4 for each group of participants. What becomes apparent is that the Complexity measures outperform the other measures no matter which sequence of target labels is used.

4.9 Conclusion and Discussion

There are three clear and novel results to be discussed:

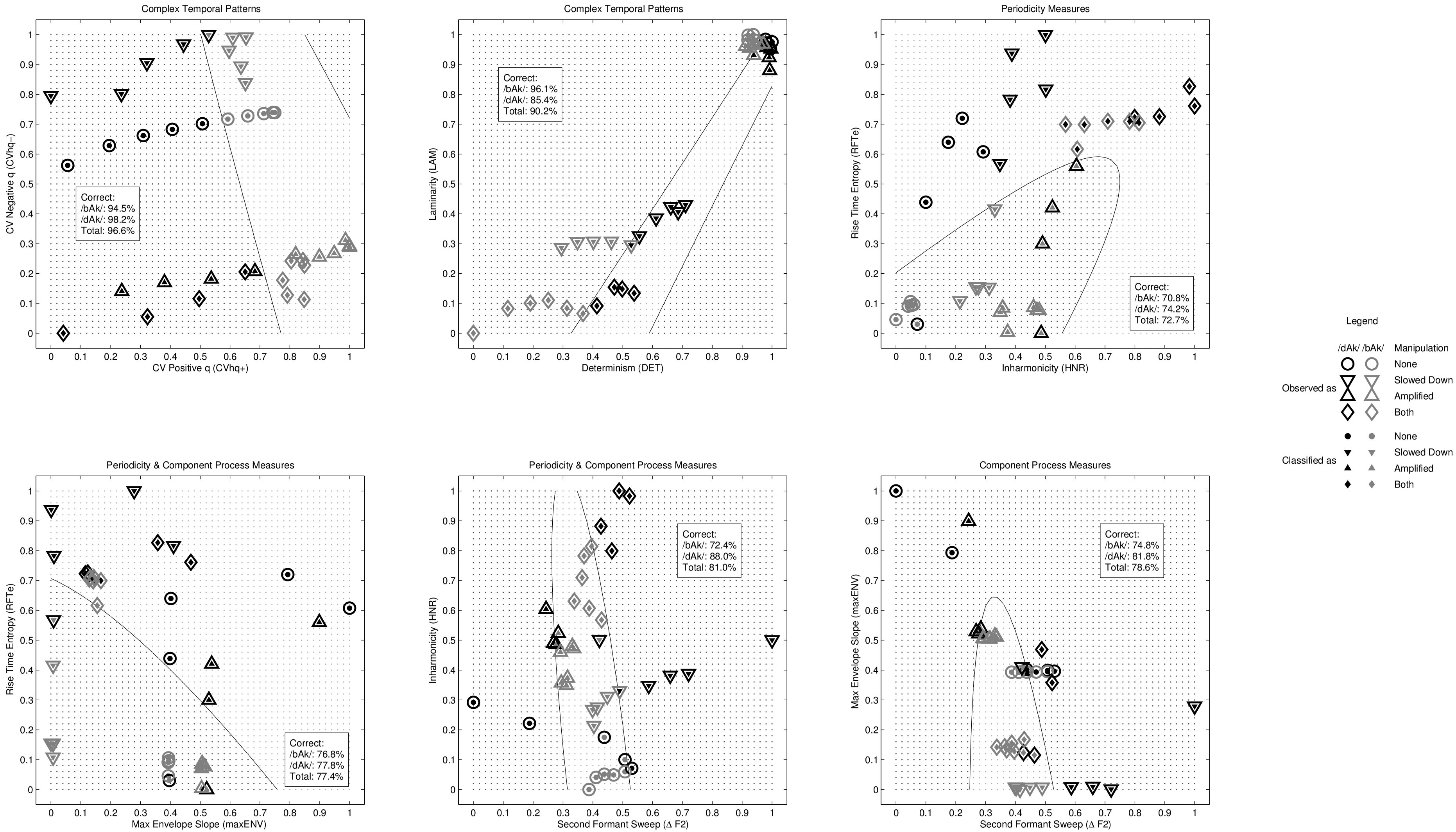


Figure 4.11 – Results of the classification of the stimuli used in the experiment by Quadratic Discriminant Analysis. Targets of the classification were the labels predicted for the sample (M_3, Table **4.6**). The panels show how the stimuli were observed (outer marker) and how they were categorised by QDA based on different pairs of measures (inner marker).

Table **4.8**
Quadratic Discriminant Analysis for Different stimulus Feature Combinations Based on Average Labelling by the Entire Sample, the Average Readers Group and the Dyslexic Readers Group. Numbers Represent Percentage Correctly Classified with 95% CI obtained from 15,000 Bootstrap Replications.

Group	Feature combination	Correct as /bAk/		Correct as /dAk/		Overall correct	
		Median	*CI.95*	*Median*	*CI.95*	*Median*	*CI.95*
Sample	CVhq+ / CVhq-	94.5%	6.2%	98.2%	8.1%	96.6%	4.5%
	LAM / DET	96.1%	10.8%	85.4%	13.6%	90.2%	6.8%
	HNR / RFTe	70.8%	19.0%	74.2%	15.3%	72.7%	8.1%
	maxENV / RFTe	76.8%	17.0%	77.8%	19.4%	77.4%	7.1%
	F2 / HNR	72.4%	17.6%	88.0%	15.2%	81.0%	9.7%
	F2 / maxENV	74.8%	18.2%	81.8%	20.9%	78.6%	8.9%
Average readers	CVhq+ / CVhq-	96.0%	8.6%	95.5%	5.2%	95.7%	4.0%
	LAM / DET	96.2%	13.4%	88.2%	13.8%	91.4%	8.8%
	HNR / RFTe	75.8%	17.8%	72.1%	11.2%	73.6%	6.2%
	maxENV / RFTe	84.3%	18.7%	81.2%	17.5%	82.4%	6.5%
	F2 / HNR	66.8%	19.4%	86.1%	16.4%	78.3%	9.7%
	F2 / maxENV	69.4%	22.2%	75.8%	25.4%	73.2%	11.3%
Dyslexic readers	CVhq+ / CVhq-	97.4%	7.1%	96.0%	6.7%	96.5%	4.3%
	LAM / DET	94.9%	15.4%	87.4%	13.6%	90.1%	9.0%
	HNR / RFTe	77.2%	19.2%	72.1%	11.9%	73.9%	4.4%
	maxENV / RFTe	84.4%	18.0%	80.9%	18.9%	82.1%	8.7%
	F2 / HNR	64.1%	17.7%	86.6%	14.4%	78.7%	9.1%
	F2 / maxENV	73.5%	20.0%	78.5%	23.1%	76.8%	11.3%

1. A difference between dyslexic and average readers in labelling some of the manipulated stimuli on the continuum is observed.

2. The Complex Dynamical Pattern measures outperform the other measures when used by a simple classifier assigning one out of two possible target labels to an observed response. This holds for the sample level as well as for each group separately, even though the sequences of target labels differ between the groups for two sets of acoustically manipulated stimuli (Amplified and Both).

3. The accuracy of stimulus classification by measures derived from different theoretical positions on the relationship between speech perception and reading appears to be ordered along a continuum (see Figure **??** and Table **4.4**). On one extreme, causal primacy is attributed to component processes (lower classification accuracy), on the other extreme, causal primacy is attributed to the interactions between component processes (higher classification accuracy).

The *first* result entails the dyslexic readers identifying stimulus 4 as /dAk/ with 95% confidence above chance when the stimulus is either amplified or slowed down and subsequently amplified. It is thus not the case that dyslexic readers "benefit" from the manipulations in terms of their speech perception becoming more like that of average readers, instead, they perceive the boundary one continuum step earlier than average readers do whenever amplification is applied to the stimuli. It should be noted though that this 'earlier' boundary perception is not the origin of the significant interaction effects between stimulus type and experimental group: the confidence intervals of the groups overlapped at these stimuli. Significant differences in odds for perceiving /dAk/ between the groups were observed for stimulus 6 (Amplified) and stimulus 5 (Both). This interaction is not likely to influence the actual labelling of the stimulus since both groups would label it /dAk/ above chance with 95% confidence. This difference would be noticed when the stimuli were presented to the same person many times in which case a dyslexic reader would label stimulus 5 (Both) about 9/10 times as /dAk/ and an average reader about 7/10 times. A similar result was found in Hasselman (2014b) where it was suggested that applying some manipulations may actually reduce the accuracy of identification and discrimination of stimuli because it biases perception towards /dAk/.

The *second* result concerns the performance of a simple classifier (QDA) employed to label the stimuli as participants in the experiment using several different measures extracted from those stimuli. The classifier performed best when the Complex Temporal Pattern measures (Coefficients of Variation of local scaling exponents of the multifractal spectrum, Determinism and Laminarity of the recurrence analysis) were used. In fact, the classification was almost perfect when the multifractal features were used. Upon examination, the only stimuli misclassified by the complexity measures were stimulus 4 (once) and 5 (six times), in both groups taken together (see Figure **4.12** and Figure **4.13**). These stimuli lie on the perceptual boundary (Stimulus 4-6) where the target label changes from 0 to 1. Misclassification may be expected for these stimuli, but classification should be relatively accurate outside of this transition region. However, this expected pattern is not what is observed for the other feature combinations. There were many additional misclassification outside the region of the label transition yielding classification curves that are clearly false (see Figure **4.12** and Figure **4.13**).

The *third* result concerns the condition of strong inference: What is the implication of these findings for the two deficit hypotheses associated with the F2 Slope / HNR measures (ATPDH) and mxENV Slope / RFTe measures (RTPDH)? First, all measures yield different values that appear to differentiate the stimuli in a sensible way (see Figures **4.2**, **4.3**, **4.4** and Table **4.2**). In other words they have the potential to be used for identification by a classifier. In fact, the classification results, expressed as % correct are not disastrous when these measures are used and at the sample level many stimuli are indeed labelled as human participants would label them. Some of these correct classifications may be expected from the way the stimuli are constructed. After all, this was done by manipulating the onset of the F2 while keeping everything else constant. Relative to that fact, their low rank in the accuracy results is surprising and should have consequences for the perceived validity of the role these features play in speech perception in general and developmental dyslexia in specific.

The measures used in this study to reveal invariant structure across scales of fluctuation, were inspired by Little and McSharry (2007) who showed RQA and scaling exponent based measures yielded the best classification of healthy and disordered speech. In such a clinical context the benefit of roughly 10% more accurate detection of disordered speech is immediately apparent. In the present study stimuli were classified, not participants and it is unlikely that the current difference in labelling between average and dyslexic readers would provide a gain in diagnostic capabilities over standardised reading tests. The difference between the groups of readers observed in Figure **4.10** are reflected in the QDA analysis by an earlier label change (at stimulus 4) for dyslexic readers labelling the Amplified and Both stimulus manipulations. The multifractal spectrum measures enable the classifier to model this early jump correctly, the RQA measures fail for the Amplified stimuli (but also in the Average Readers group). The other measures fail for both stimulus types producing earlier jumps (Stimulus 3 or earlier) or later jumps (Stimulus 5 or later) in dyslexic readers, for average readers these patterns are shifted up the continuum (Stimulus 4 or earlier and Stimulus 6 or later). Apparently, there are invariant temporal structures in all the audio files that are insensitive to any disruption (e.g., the acoustic manipulations) or absolute differences in physical characteristics associated with articulatory cues (e.g. due to the changing F2 onset): Their relative rank order on the labelling curve remains approximately the same.

Recent studies in speech signal analysis and animal vocalizations have indeed shown the frequency domain obtained by Fourier decomposition may not be the information used by the neural systems of mammals to perceive sounds, whereas the Hilbert decomposition in slow varying envelope and fast varying fine time structure (the analytic signal), may be the more likely candidate (Smith et al., 2002). The Rise-Time Perception Deficit Hypothesis of dyslexia (cf. Goswami

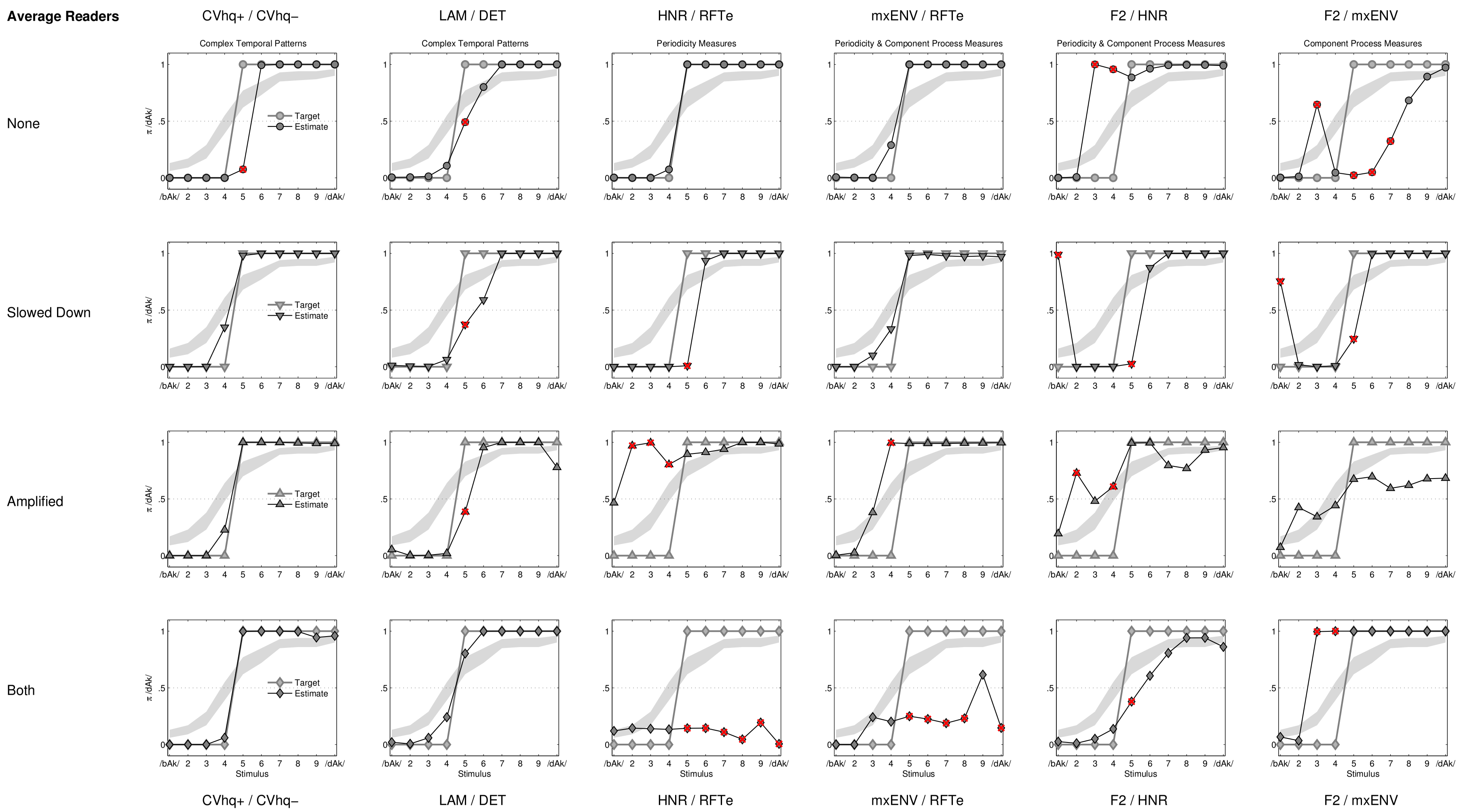


Figure 4.12 – Figures represent for each feature combination (columns) and manipulation (rows) the QDA estimated class membership probabilities (black lines), the target label for classification (grey lines) and the confidence band (grey area) predicted for Average Readers by model M_4 (see Figure **4.10** and Table **4.7**). If the lower confidence band crosses the 0.5 threshold the target label changes from /bAk/ to /dAk/. The red crosses mark misclassified stimuli.

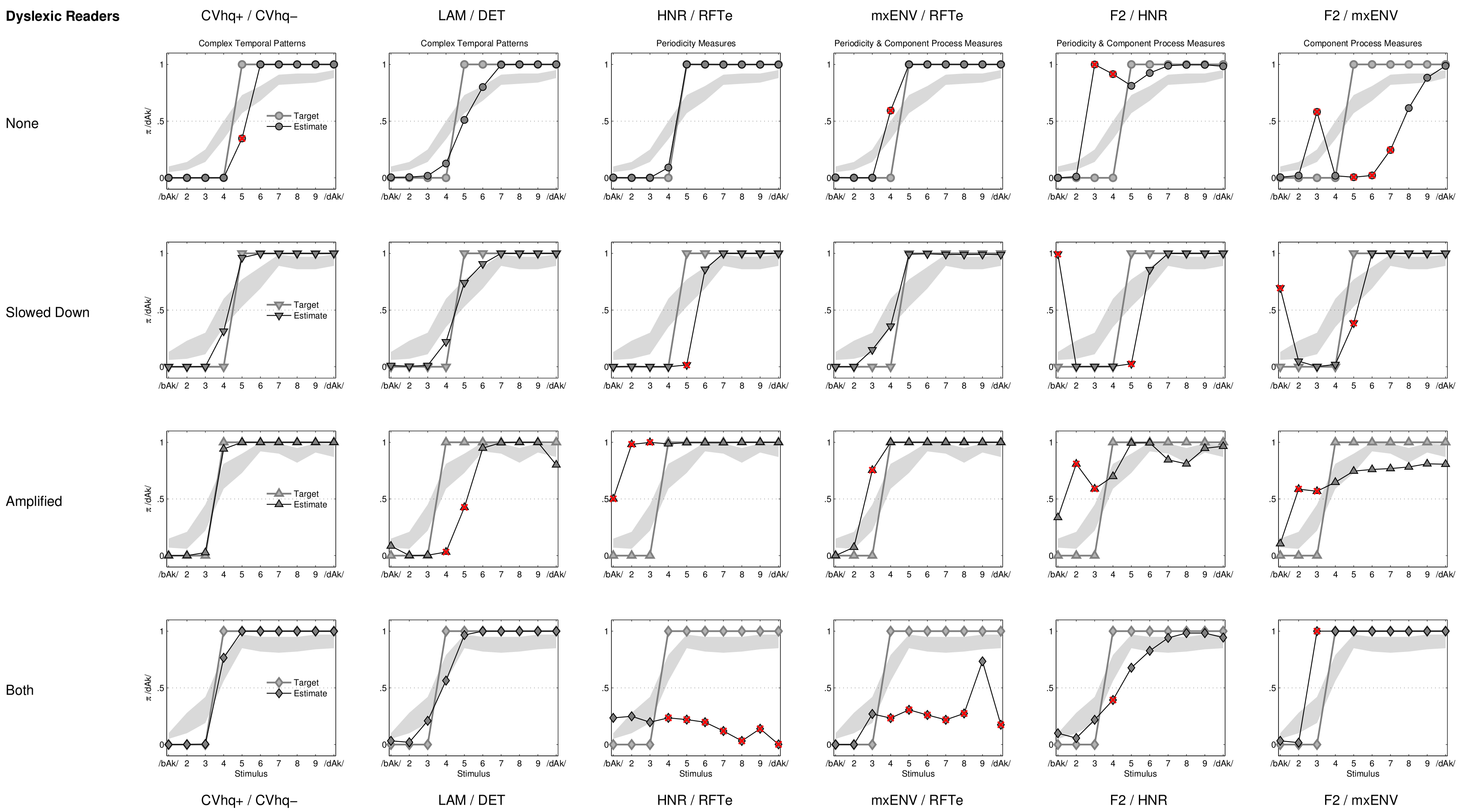


Figure 4.13 – Figures represent for each feature combination (columns) and manipulation (rows) the QDA estimated class membership probabilities (black lines), the target label for classification (grey lines) and the confidence band (grey area) predicted for Dyslexic Readers by model M_4 (see Figure **4.10** and Table **4.7**). If the lower confidence band crosses the 0.5 threshold the target label changes from /bAk/ to /dAk/. The red crosses mark misclassified stimuli.

et al., 2002) is partially based on these findings. However, the fact that the speech signal is the product (i.e., multiplicative temporal interactions) of the fast analytic signal and the slow changing envelope is not considered by the theory. In any case, the claim that speech sounds are being stored in memory as strings of abstract symbols that represent speech components such as formants and phonemes, becomes untenable when they are directly compared to features that quantify dynamical invariants presents in the signal (see Port, 2007, for a review of arguments against positing 'phone' components). Many of the traditional problems with the scientific explanation of speech perception and production appear to be related to the use of a causal ontology that posits independent components whose additive interactions generate complex behaviour such as communication by means of spoken language.

The claim is not that humans use a neurological equivalent of QDA to identify speech sounds, the present study shows that it is very unlikely that participants simply analyse (relative) frequency changes or amplitude envelopes and somehow match them to collections of frequencies and amplitude patterns stored in the brain. It also seems unlikely that a failure to match those stored features can constitute an aetiology for observed reading and spelling problems in developmental dyslexia. Instead, based on the complexity measures QDA assigns a correct classification curve to each experimental group, even when the curves differ between the groups. Compared to average readers, the category switches are 'early' for dyslexic readers which could indicate a lower threshold for perceiving /dAk/ or an enhanced contrast (see e.g., Case et al., 1995; Tuller et al., 1994) compared to the average readers. A comparison of the classification curves in Figure **4.12** and Figure **4.13** reveals that the multifractal and RQA measures which does not appear in any systematic way for other measures. This suggests that the processes underlying the small observed labelling differences between average and dyslexic readers may indeed reflect a scaled continuum rather than a specific impairment, a deficient component.

4.9.1 The classical information processing problems: Lack of invariance?

Recently, Kleinschmidt and Jaeger, 2014 described an 'ideal adapter framework' based on (Bayesian) belief updating to model three challenging aspects of speech perception: 1) *Recognize the familiar*, 2) *Generalize to the similar*, and 3) *Adapt to the novel* (Kleinschmidt and Jaeger, 2014, p.4). These well known problems in the scientific study of speech perception are related to the *lack of invariance* between speech signals that are perceived to be similar, when in fact they differ substantially with respect to one or more physical characteristics of the produced signal (Liberman et al., 1967). The F2 manipulation in combination with the acoustic manipulations applied in this article can be considered a modest example of such variants, in reality the differences between speakers in the production of an F2 onset may be much more extreme than represented by the stimulus set used in the current study (see e.g., Kleinschmidt and Jaeger, 2014). The *similarity recognition problems* (point 1 and 2 above) emerge due to the conception of perception and recognition memory as a database search prompted by an 'incoming' query (the signal). Specific values of perceptual cues are hypothesized to lay dormant, stored inside the brain, waiting to be constructed into a larger whole by accumulating matching stimulus features. Due to the lack of invariance, these features must somehow be collected into aggregate sets of features that overlap considerably between different categories.

To illustrate how the similarity recognition problem arises from its conception of a search and match operation, consider the mechanism behind a popular application for smartphones called Shazam (Wang and Chen, 2003). It is capable of analysing music being played in the environment and after a few seconds it provides the name of the song and the artist who performed it. Interesting features are that it does not matter which part of the song is analysed and that as long as the recording being played exceeds background noise and is in the Shazam database, a few seconds of analysis are enough to yield almost 100% accuracy. The search and match time is reported to be between 5-500 milliseconds. Based on a sound recording a unique time-coded fingerprint is extracted from the spectrum and is stored in a database. If a song needs to be recognised, a smart search algorithm can quickly find likely candidates for the origin of the small sample of the fingerprint (Wang and Chen, 2003). The fingerprints are so unique any song in the database can be quickly identified, irrespective of the sample being taken from the begin, middle or end of the song. This is exactly the reason why the database query metaphor is an unlikely model for speech perception: Humans are generally not very good at accurately reconstructing a word or sentence when just one or two parts (phones, words) are presented. The requirement of uniqueness in this type of database search is the main cause of the apparent similarity recognition problem in speech perception. A song of which the original studio recording is in the database will not be recognised when a sample of a live recording of the same song performed by the same artist is the source of the query. This is a failure to recognize the familiar, because the system cannot generalize to the similar. Even a studio recording of the same song by the same artist, but with a different audio mix (e.g., older recordings that were 'remastered') will not be recognised if the actual recording is not stored in the database.

This problem of generalisation is one of many problems identified with the notion of perception as constructing meaningful information from incoming perceptual cues by matching it to stored meaningful information (see e.g., Chemero, 2009; Haselager et al., 2003). Even if one wants to propose that we just store everything we hear from the day we are capable of doing so and disregard the fact that the amount of meaningful information to be stored would become infinitely large, it means we cannot understand someone the first time we meet him or her. We first have to store into a database the fingerprint of his or her utterances, using different speaking voices! Merleau-Ponty described it as follows: "*An impression can never by itself be associated with another impression. Nor has it the power to arouse others. It does so only provided that it is already understood in the light of the past experience in which it co-existed with those which we are concerned to arouse.*" (Merleau-Ponty, 1962, p.14). The internal representation of experienced reality is an unnecessary assumption in understanding intelligent behaviour when one examines how human perception and action is constrained by the physical features of the body and the environment (Dreyfus, 2002).

Biological information processing: Abundant self-affine invariance?

The interaction dynamics that give rise to a constraining of the degrees of freedom in human speech perception and production were lucidly described by Stetson: "*Speech is rather a set of movements made audible than a set of sounds produced by movements*" 1951, p.33. So the 'set of features' that should reveal the invariance used in categorical perception should be related to the complex system that produces the speech signal. There is evidence that a close bi-directional perception-action coupling exists when speech perception and production are concerned. In a series of experiments Perkell et al. (2004b,a) have shown that the distinctness, or quality of a produced vowel contrast by a speaker, is related to the quality of the perception of that contrast by the same speaker. In other words, speech production will constrain speech perception and vice versa. Some of these notions have been incorporated in the DIVA (Directions Into Velocities of Articulators) model of speech production (Guenther and Perkell, 2004). In short, this model learns to produce speech by tuning, or constraining its motor output to auditory targets it is presented with (like an infant would attune to the often very repetitive speech-like utterances produced by its parents). This is in principle the same 'mechanism' suggested by the complexity matching hypothesis.

In the present context of self-affine scaling, the recognition of familiarity and generalization to similarity are represented by the different scaling relations estimated to constitute the spectrum of generalized Hurst exponents. That is, the local scaling exponents quantify the magnitude of 'familiar similarity' (right part of Figure **4.9**) relative to the signal *itself*, observed at different scales of

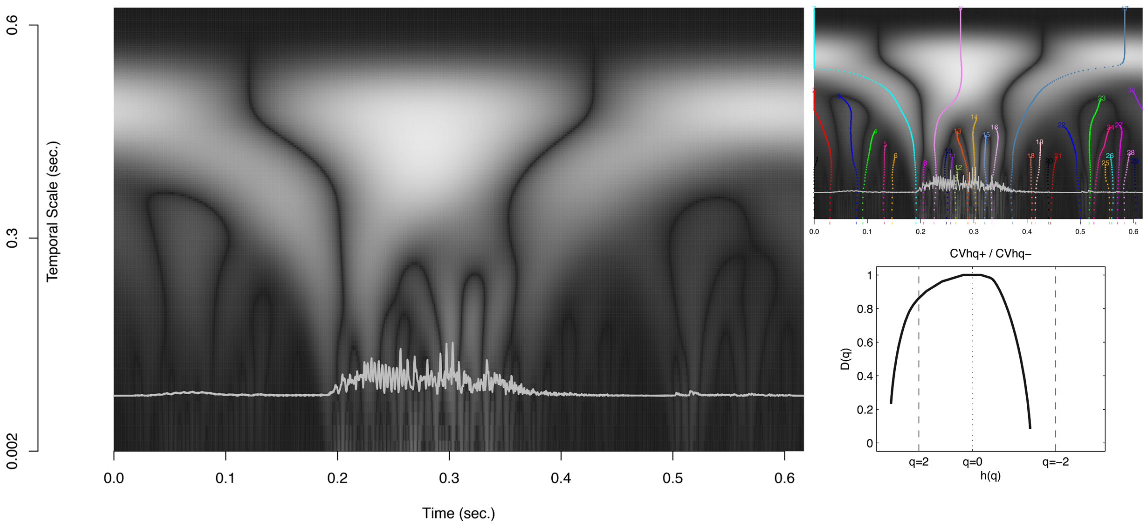


Figure 4.14 – A scaleogram representing a Continuous Wavelet Transform of the amplitude envelope of Stimulus 1 (grey line). On the right, the top figure displays the wavelet singularity extrema as coloured traces that connect the different scales at which the wavelet is associated. and the singularity spectrum. See text for details.

fluctuation (left part of Figure **4.9**). Figure **4.14** reveals the full multi-scale, self-affine structure of temporal patterns present in the signal by means of a Continuous Wavelet Transform of the signal. The x-axis in the scaleogram represents time and the y-axis represents scales of fluctuation (expressed in seconds). The colour-coding represents the goodness-of-fit of the shape of continuously scaled versions of a 'mother' wavelet (the Mexican hat) with the shape of the observed signal. The scaled shape is shifted across the time axis and this causes the change in colour from left to right. The process is repeated for different scaled versions of the wavelet and this causes the change in colour from top to bottom. If the wavelet is scaled to cover large portions of the time-series, the fluctuation frequencies it can detect will be slow fluctuations and vice versa. In Figure **4.14** the largest scale is about 0.6 seconds and the dark colour indicates the expected low association between the stretched wavelet with the entire signal. The large light coloured branching structures that extend across many scales reveal how patterns recorded at the smallest scales are nested as self-affine scaled copies within the larger structures.

By following the vertical extrema of cross-scale associations (the vertical line structures), so called temporal singularities can be found, that occur when the structure at a larger scales branches into two smaller vertical structures (see Figure **4.14**, inset on the right). These singularities constitute a spectrum that is equal to the generalized Hurst spectrum. The coloured lines in Figure **4.14** trace a path that provides information about the signal that is invariant across many scales. Some paths yield predictive information (a larger scale version of the current waveform is yet to come), others constrain (or confirm) what has already occurred (the current waveform is a scaled version of larger wave form that just occurred). The entire spectrum can be considered a complex resonance frequency for self-affine structure. The adaptation to novelty achieved by QDA (i.e., adaptation of the classification solution based on slightly different empirical curves) is 'simple' enough to consider physically realizable in a biological system. A self-tuning resonator (Collins et al., 1995; Gammaitoni, 1995) could be an interesting metaphor.

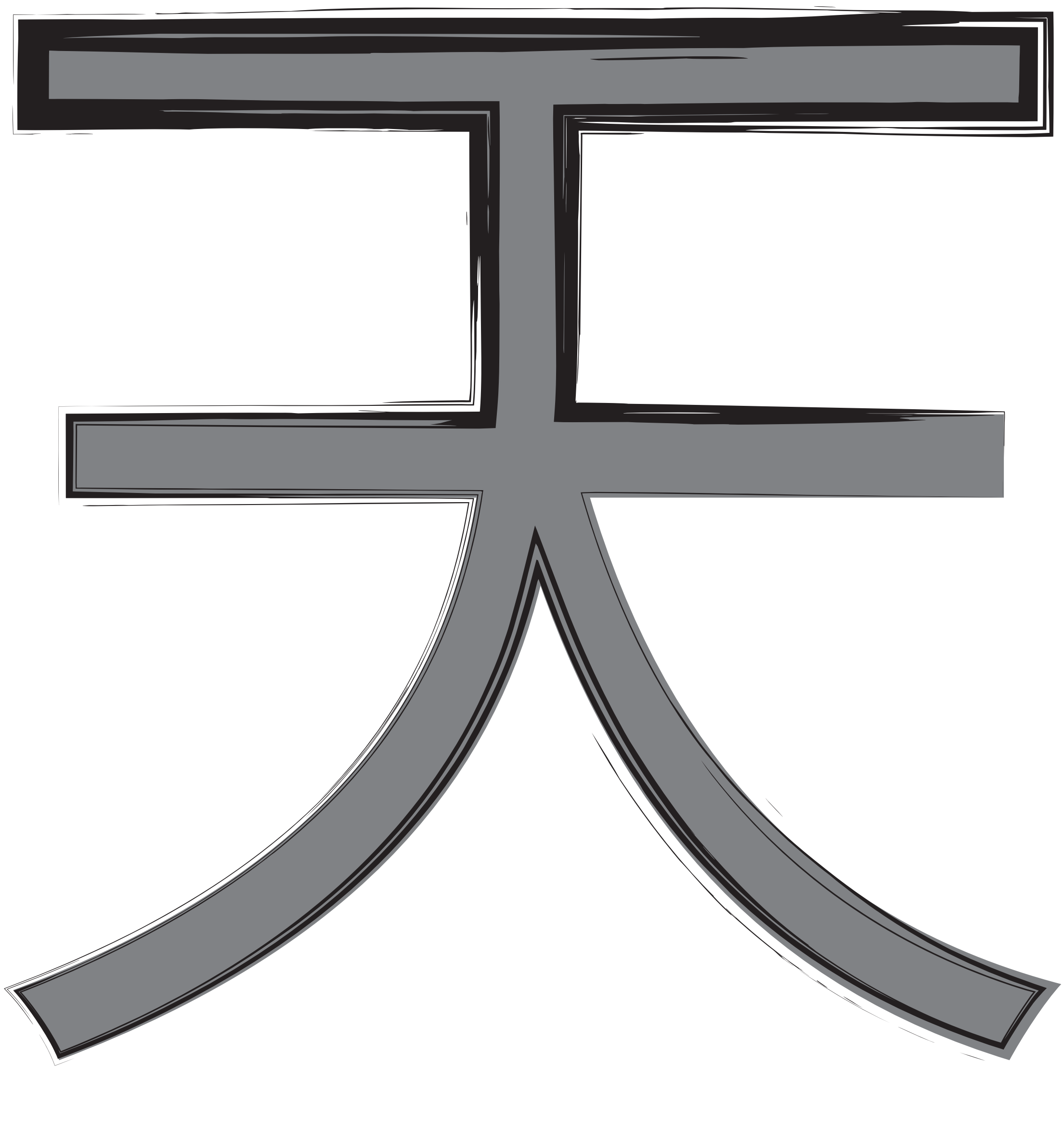
4.9.2 General Conclusion

Whether participants actually matched, or resonated the complex dynamical pattern remains a topic of future studies: To evidence such matching at the scale at which the speech sound unfolds would require (neuro-)physiological measurements. The global convergence of the classifier accuracy on the scale of component-dominant to interaction-dominant causal ontologies of behaviour is non-trivial. The former perspective looks for components as efficient causes of behaviour (e.g. % of variance in one variable that is uniquely attributable to the levels of another variable) whereas the latter looks for dynamical invariants and correlations across lags time that may be exploited to coordinate behaviour (e.g.,long range anti-persistent correlations or self-organised critical states, cf. Van Orden et al., 2003). Although it is important to note that this does not mean an interaction-dominant perspective denies that components exist, it does imply that components (from phoneme representations to 'cues') should be assigned a different causal role in production and perception of human speech. It follows that components and component processes proposed by RTDH and ATPDH should be reconsidered as a factor in the aetiology of developmental dyslexia. The current results do not provide a readily available alternative, but they do provide strong cause for the development of an aetiology based on an interaction-dominant causal ontology, for example based on the scaled continuum hypothesis (Hasselman, 2014a; Holden et al., 2014; Wijnants et al., 2012b) and complexity matching.

It is of course important to replicate these findings with other stimuli and other samples of participants. Interestingly, the analysis presented here can be performed post-hoc on any speech identification study already published. The measures can be extracted from any signal and the QDA can be applied using the observed labels found in the study as targets for the classification.

4.10 Acknowledgements

I would like to thank Anna Bosman and Ludo Verhoeven for their comments on a previous draft of this article. I am also very grateful to the anonymous reviewer who suggested inclusion of (multi)fractal measures would provide a stronger case for the Complexity Matching Hypothesis.


天

"What I call the void is where nothing exists. It is about things outside man's knowledge.
Of course the void does not exist.
[Of course the void is nothingness]
By knowing what exists, you can know that which does not exist.
That is the void."

- Shinmen Musashi (1645, The Book of the Void)

In the fifth chapter of "A Book of Five Rings", completed just weeks before his death, the legendary Samurai Miyamoto Musashi explains the inexplicable.

Apocrypha

Miyamoto, Musashi (1974). A Book of Five Rings, translated by Victor Harris. London: Allison & Busby; Woodstock, New York: The Overlook Press.

Notes

Chapter 5: Principled Simulation & Strong Inference

- This chapter may be cited as:

  Hasselman, F. (2014). Beyond the Static Phoneme Boundary: The Nonlinear Dynamics of Emerging Literacy. In *Beyond the boundary. An analysis of verisimilitude and causal ontology of scientific claims: Ætiologies of developmental dyslexia as a case in point*. fig**share**. http://dx.doi.org/10.6084/m9.figshare.1080795

- The supplemental materials to this chapter may be cited as:

  Hasselman, F., (2014). Beyond The Boundary - 5th Chapter: Supplemental Materials. Retrieved from Open Science Framework, https://osf.io/v3wjn

Chapter 5

Beyond the Static Phoneme Boundary

The Nonlinear Dynamics of Emerging Literacy

A common misconception about learning to read is that it requires of the aspiring literate to be aware of spoken words being composed of smaller sound units called phonemes (Port & Leary, 2005; Port, 2007). It is indeed the case that preliterate children and illiterate adults can often segment words into onsets and rhymes, but manipulating individual speech sounds like identifying, adding or omitting a sound turns out to be a skill very few posses before receiving some literacy training (de Graaff, Hasselman, Bosman, & Verhoeven, 2008; de Graaff, Hasselman, Verhoeven, & Bosman, 2011; Goswami, Ziegler, Dalton, & Schneider, 2003; Morais, Cary, Alegria, & Bertelson, 1979; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). This is a phenomenon that is observed across languages, but especially in languages with alphabetic writing systems. In fact, one could argue that the phoneme is a cultural phenomenon rather than a perceptual reality of body and brain (Port, 2010a). Even in an almost entirely transparent language (i.e. the spelling of a word maps almost exactly to the way it is pronounced) like Finnish there is no pre-literate phoneme awareness. In Finland formal reading instruction starts quite late compared to other countries, a recent study revealed that the most basic phonological awareness and early reading skills develop simultaneously in that language (Silvén, Poskiparta, Niemi, & Voeten, 2007). That is, exposure to the alphabetic writing system was likely the driving force behind the ability to segregate speech into phonemes. The authors concluded that language- and culture- specific influences on literacy development must be considered in models and theories instead of assuming innate perceptual boundaries.

5.1 Phoneme Awareness & Emerging Literacy: Two Sides of the Same Coin?

The component dominant dynamics accounts of developmental dyslexia leans heavily on the assumption that phonological deficits are the core deficit in dyslexic readers (Landerl & Wimmer, 2000). This might be consistent with results from studies done in the English language (a rather opaque orthography), but seems to be only partly true for languages with more transparent orthographies. In German for instance phonological abilities seem to play only a marginal role in (predicting) developmental dyslexia (Wimmer & Schurz, 2010). Goswami (2000) and Vellutino, Fletcher, Snowling and Scanlon (2004) reviewed several prospective studies conducted in different countries and both studies concluded that for instance rhyming skill is a good predictor of reading difficulties in English, but not in German and Dutch. In countries with transparent orthographies, rapid automatised naming (RAN) seems to explain more variance than phonological ability. A study on the precursors of literacy in Finnish by Puolakanaho et al. (2007) reports that measurements of letter naming, phonological awareness tasks and rapid automatized naming among kindergartners taken together correctly categorize 80% of children who turn out to develop reading problems in 2nd grade. Similar results have been reported in prospective studies on the acquisition of reading in Dutch (Eleveld, 2005; van den Bos, lutje Spelberg, & Eleveld, 2004). The exact role of rapid naming in literacy development remains a matter of debate (Moll, Fussenegger, Willburger, & Landerl, 2009; Vaessen, Gerretsen, & Blomert, 2009). The point here is that the assumption that learning to read requires awareness of phonemes before the connections between letters and sounds can be learned is too simple. The strong claim I will defend in this chapter is that awareness of phonemes is caused by exposure to an alphabetic system of symbols used to represent (a small part of) the sounds used by speakers of a language (Linell, 2005; Port, 2007, 2010a, 2010b).

It has been repeatedly shown that systematic phonics training for kindergarteners increases their phonological awareness (e.g. in Dutch: de Graaff, Bosman, Hasselman, & Verhoeven, 2009; de Graaff, Verhoeven, Bosman, & Hasselman, 2007), also in children at risk for dyslexia and speech-language impairment (Segers, Hasselman, Verhoeven, & De Graaff, 2004). Even if an adult poor reader has received only rudimentary phonics instruction in childhood, this is noticeable as better performance in phonemic awareness tasks when compared to illiterate peers (Morais et al., 1979). A crucial finding that shows the impact of exposure to an alphabetic writing system involved two groups of Chinese adults. One group was literate in characters and alphabetic spelling (Hanyu pinyin) and could manipulate consonants by adding or omitting them from Chinese spoken words. A group of adults literate only in characters could not perform these simple speech-sound manipulations (Read, Zhang, Nie, & Ding, 1986).

This is a problem for the role of the deficient phoneme representation as a cause of developmental dyslexia. If there is no pre-literate sound representation at the level of the phoneme, its impaired-ness or non-distinctiveness (e.g. Elbro, Borstrøm, & Petersen, 1998; Snowling, 2001) cannot be the explanation behind the difficulty of acquiring grapheme to phoneme correspondence rules. The only thing one could argue for is that there must be some process that prevents the representation of phonemes from being formed properly during the time in which children receive literacy training in an alphabetic writing system. If the radical position is taken seriously and one wants to dispense with abstract phonological codes that refer to segments of speech, the causes of the reading impairment must be sought in other acoustic features of the speech signal that allow segmentation, categorisation, and association to the categories of written language (Goswami, Fosker, Huss, Mead, & Szűcs, 2010; Ramus, Nespor, & Mehler, 2000).

5.2 A Speech Perception Deficit Without Impaired Representations

What are the consequences of taking the radical position on the reality of phoneme representations for the speech perception deficit hypotheses of dyslexia? A logical deduction based on the impaired phoneme representation hypothesis is that the categorisation and discrimination of speech sounds by dyslexic readers should be impaired (Bradley & Bryant, 1983). If phoneme representations or the awareness of sound segments indeed emerged only during or after exposure to an alphabetic writing system, could there be a literacy exposure effect on perception of speech sounds? There are some studies that certainly seem to point in that direction; a comparison of dyslexic readers and illiterate adults for instance seems to point to similar deviations in classification curves of those groups compared to typical readers (Serniclaes, Ventura, Morais, & Kolinsky, 2005). Another study included reading age and chronological age control groups in a comparison of speech perception ability with a group of dyslexic readers and observed a literacy exposure effect on speech perception. The authors concluded however that discrimination curves of dyslexic readers could not be fully explained due to a lack of reading experience alone (Bogliotti, Serniclaes, Messaoud-Galusi, & Sprenger-Charolles, 2008).

Many of the arguments against a mental alphabet of discrete speech sound representations are based on the lack of invariance and discreteness of the acoustic features supposedly associated with these abstract representations. Speech is a continuous signal not discrete; speech gestures overlap (co-articulation) and acoustic features like formants are perceived depending on the context of other features present in the signal. In other words, speech perception and production are highly context sensitive. The variations in the hypothesised speech units like phones and phonemes produced by different speakers with respect to pitch, voice onset, and formant frequencies are almost infinite (Hay & Drager, 2007 is a review). Concerns about the nature and prevalence of a speech perception deficit in developmental dyslexia have their origin in such lack of invariance across contexts, that is context sensitivity of study outcomes (e.g. Serniclaes, 2006). As was shown in Chapter 2 there are obvious and less-obvious contextual factors that appear to influence whether a deficit in speech perception is found. In Chapter 4, I suggested that the higher order dynamics of the speech signal might be the invariant feature humans use to identify speech sounds. In the literature many other factors are reported that influence whether a speech perception deficit is found; the nature of the stimuli used in experiments (natural vs. synthetic speech: Blomert & Mitterer, 2004), whether a specific subgroup of dyslexic readers is examined (deficit found mainly in language impaired dyslexic readers: Manis & Keating, 2005), or specific task constraints (impaired phoneme representations only found under time pressure: Ramus & Szenkovits, 2008).

Causation, everywhere!

As mentioned in previous chapters, there is also much debate about the underlying causes of the impaired phoneme representations (Ramus, 2003, 2004), there are two main views both backed by neuroanatomical evidence that differ in the localisation of the cause, being peripheral (perceptual) or cognitive (language related). The peripheral view states that the impairment is caused by a deficit in low-level auditory processes caused by magnocellular disruptions in the thalamus and the cerebellum (see for instance Farmer & Klein, 1995; Stein, 2001; Stein & Walsh, 1997). The cognitive view states that the cause is speech and language specific (Mody, Studdert-Kennedy, & Brady, 1997) and is caused by neuroanatomical anomalies in the perisylvian cortex (Ramus, 2004). Recent evidence suggests this apparent clear-cut distinction between auditory and 'speech-language' accounts that are each backed by considerable empirical record, may prove to be the result of a subtle instruction effect. When the same set of sine wave stimuli are in one occasion presented as being electronic whistling sounds that need to be discriminated and later as speech-like sounds, the second instruction causes participants to perceive a phoneme boundary that is not observed when the first instruction is given (Serniclaes, Sprenger-Charolles, Carre, & Demonet, 2001).

This effect has been interpreted as a so-called 'speech mode' in which a non-speech stimulus is interpreted as a speech sound. The speech mode has been shown to have neural correlates (Dehaene-Lambertz et al., 2005). Several experiments are reported in the brain-imaging study that basically uses the same instruction variation as mentioned above, adapted for measurements using EEG, MEG, and fMRI. The authors report three main conclusions, of which the first is a confirmation of the instruction effect found by Serniclaes et al. (2001). As an example of the problems that arise when these subtle contextual effects are interpreted in terms of discrete abstract representations of invariant acoustic features, I take a moment to analyse in some detail the way the authors interpret their results:

> ``First, the same auditory stimuli are processed differentially depending on whether they are perceived as speech or as nonsense electronic whistles. Second, the posterior part of the superior temporal sulcus and the supramarginal gyrus are crucial areas for syllable processing but are not involved in the processing of the same physical dimension when the stimuli are not perceived as speech. Third, non-phonemic auditory representation and phonemic representation are computed in parallel, but the phonemic network is more efficient and its activation may have an inhibitory effect on the acoustical network." (Dehaene-Lambertz et al., 2005, pp. 32).

The first conclusion seems straightforward, but is circular: The same stimuli are processed differently depending on how they are perceived. How does the brain of a speaker know whether a sound will be perceived as a tone or as speech before it is even processed? Of course, in the context of the experiment the participant is primed by instruction, but how does this generalise to real life situations? The same problem applies to the second conclusion: How are the brain regions that process the signal informed by the percept? That is, isn't the purpose of the entire perceptual process to figure out what the acoustic pattern may represent in the first place? As a pure description of the experimental findings these conclusions might be acceptable when interpreted as reporting a correlation: This is what we observe when a stimulus is perceived as speech rather than as a sound. The circularity emerges due to the suggestion that different brain regions are involved in the processing of the same physical dimension of an auditory signal dependant on the outcome of this process. That's time travel.

The third conclusion seems to confirm the assumption that the authors are not just reporting a correlation, but are interpreting the difference due to instruction they observed as the causal power of the phonemic representation to suppress an acoustic pattern to be registered as an auditory percept (before the perceptual process is completed). They also inflate the amount of representing being done by the brain. Apparently the same auditory stimulus has two different representations, one is non-phonemic and is processed by the auditory network, whereas the other is phonemic and processed by the phonemic network. The consequence must be that all auditory stimuli of which the physical dimensions may be perceived as non-speech as well as speech -but are not speech sounds- take up storage space as a speech-sound representation as well as an auditory representation in the brain. It is almost impossible to break out of the circular reasoning: When a non-speech stimulus is perceived as a speech sound, the phonemic network, which is much more efficient, suppresses the auditory network causing a speech sound to be perceived.

Note that this problem arises due to the assumption that simple linear efficient causes independently add up to coordinate behaviour. Each qualitatively different outcome must have a different composition of causes attached to it. In this chapter I will discuss a model that allows an alternative interpretation of these and similar results in which the (phoneme) representation is invoked into existence as an explanatory vehicle that is suggested to be identifiable as a physiological trace or a structural component of the nervous system. The only thing that is in fact observed on the level of behaviour is a transition of one stable perceptual state into another. To introduce the model I will first discuss an interesting hypothesis recently suggested to explain some surprising characteristics of speech perception by dyslexic readers. No surprise, a new neurophysiological component representation of patterns in the acoustic signal is invoked into existence as to explanation the observed behaviour.

5.3 The Elementary Particles of Speech Sounds: Phones, Allophones and Phonemes

A remarkable and quite recent finding is that dyslexic readers outperform average readers in within category discrimination of speech sounds (Serniclaes, Van Heghe, Mousty, Carré, & Sprenger-Charolles, 2004; Serniclaes, 2006; Werker & Tees, 1987). The explanation is that in dyslexic readers rudimentary pre-linguistic phoneme categories, or allophones, still exist as separate categories along a continuum. An allophone is a speech sound or phone that is not perceived as a distinct speech sound or phoneme in a particular language by adult listeners. Thus, a phoneme can be a collection of allophones or similar sounding speech sounds that are categorised under one label, the label of the phoneme. Exchanging one allophone for another in a word may sound awkward, but does not change it's meaning. Whether a phone is an allophone may differ for different languages.

To illustrate this, the voice onset time (VOT) continuum from /d/ to /th/ is divided into two categories in for example, French, Spanish and Dutch whereas in Thai it is divided into three categories (Abramson & Lisker, 1970; Lisker & Abramson, 1971; Serniclaes, 2006). This categorization of the different allophones along the VOT continuum into three phoneme categories in the Thai language is considered the “natural” division and is thought to arise from a predisposition to categorise certain features into categories by pre-linguistic children (Lasky, Syrdal-Lasky, & Klein, 1975; Serniclaes, 2006; Werker & Tees, 1984). Hence, in three category languages the perception of phoneme boundaries along the voicing continuum in pre-linguistic children will be roughly the same as in adults. In two-category languages a child has to merge or couple allophones that naturally would be perceived as three distinct phonemes into two distinct phoneme categories; this is called the coupling hypothesis (Serniclaes, 2006). Speech perception of adults in such languages differs from that of pre-linguistic children, at least in the first year of life (Werker & Tees, 1984). Dyslexic readers in such a language, according to this account, have failed to successfully couple these voicing allophones to the two phoneme categories used in their native language and their speech perception will resemble that of the pre-linguistic child. Their better discrimination ability of stimuli that vary along a specific acoustic continuum but within the same phoneme category is thus considered to be due to an allophonic mode of speech perception (Serniclaes, 2006; Serniclaes et al., 2004). Serniclaes and colleagues note that these pre-linguistic boundaries can still be found in non-dyslexic readers, but to a much lesser extent.

5.3.1 Beyond Explanatory Boundary?

Here we encounter the limits of the theory of elementary speech particles. Why do pre-linguistic children categorise a certain physical continuum related to speech sounds into three categories at specific values of that continuum? “Because they were created to do so”, is the uncomfortable translation of predisposition in this context. If we were indeed created to do so and evolution spent a lot of time and energy to encode these boundaries into our genes, then why are not all the languages of the world using these innate boundaries to segment speech? And is not it strange that Thai appears to be the only language that uses these predisposed category divisions? Careful examination of the older studies upon which much of the allophone perception argument is based, reveals that the data are not at all consistent in their support for the existence of exact universal phoneme boundaries in infants. Nor do they point at the existence of exact acquired boundaries in adults.

Three key assumptions of the allophone perception hypothesis can be disputed based on studies that go back some 50 years:

1. Adults cannot discriminate between speech stimuli within a native phoneme category.
2. Infants perceive speech categorically, indicating innate phone-like representations more fine-grained than phoneme contrasts of a language. This is why infants can discriminate between all non-native language phoneme contrasts.
3. Speech perception, in general, is categorical, not continuous.

Assumption 1. Adults can’t discriminate within categories [incorrect]

Concerning the first assumption, Abramson and Lisker (1970) reported that the pooled discrimination curves of a VOT continuum by English speakers showed one discrimination peak and the curve of Thai speakers indeed showed two peaks (indicating three categories). There were however many (consistent) individual differences between speakers of the same language, who perceived the boundaries at completely different VOT values from other native speakers (Abramson & Lisker, 1970). Also, when the production of VOTs is studied for different languages as reported by Lisker and Abramson (1971), there is much variability in the exact location of the produced VOT values along the continuum. Abramson and Lisker (1973) indeed found Spanish phoneme boundaries at VOT values that were expected based on production of word-initial stops in Spanish. That is, on the aggregate level. There were many individual differences in produced VOT. Surprisingly, and contrary to claim 1, adult Spanish speakers were able to discriminate between stimuli within a native Spanish phoneme category.

Based on these and other observations Lisker and Abramson (1971) argued in favour of the continuous control of voice onset time by speakers and rejected the notion by Chomsky and Halle (1968) that arrangements of speech segments with fixed feature values are able to explain their VOT timing data. Continuous, adaptive control instead of a pre-programmed set of gestures is indeed what studies suggest that examined effects of the perturbation of the jaw, lips, and tongue during speech production (Houde & Jordan, 1998; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Munhall, Löfqvist, & Kelso, 1994; Nittrouer, Munhall, Kelso, Tuller, & Harris, 1988). Results show that speakers recover within milliseconds after unexpected movements of the jaw, lips or tongue caused by a mechanical device specifically created to disrupt the speech production of participants. Such ultrafast recovery cannot be explained by ballistic execution of motor programs (cf., Van Orden, Hollis & Wallott, 2012).

Assumption 2. Innate phones develop into categories [incorrect]

The second assumption concerns speech perception by infants, who, according to a well-known science fact are supposed to be able to discriminate between all the speech sound categories of the languages of the world, up to a certain age. Some older studies are often quoted to support categorical perception by infants as young as two months old. In one case this turns out to be discontinuous perception (perceptual change or identification of sounds along a continuum without clear category boundaries) rather than exact categorical perception (Eimas & Miller, 1980). As a whole, the actual empirical evidence is more complex than the popular fact implies: Spanish 6-8 month old infants were able to discriminate boundaries on both Spanish and English continua, but English infants of the same age only perceived the boundary native adult English speakers perceive (Eilers, Gavin, & Wilson, 1979). Much like the contemporary discussion about auditory versus speech specific perception, studies have been published showing 2 month old infants can also perceive non-speech sounds as categories, casting doubts on the special perceptual status of speech sounds and speech sound categories (Jusczyk, Rosner, & Cutting, 1977). More recent results show that infants are indeed able to discriminate between stimuli within an adult phoneme category (McMurray & Aslin, 2005). The second claim should not be seen as a corroborated fact.

Assumption 3. Perception is discrete [incorrect]

The final assumption, categorical versus continuous perception has been, and still is a subject of many studies. It has been known for some time that there are important features of the speech signal that are used for identification of sounds, but that do not yield a discrimination peak when placed on a continuum, such as the non-categorical perception of tonal categories in Thai (Abramson, 1979). Non-categorical perception of speech sounds can also be induced by experimental manipulation (Carney, Widin, & Viemeister, 1977; Pisoni & Lazarus, 1974). In a study where subjects were asked to give continuous rating responses about how well they thought a speech stimulus belonged to a category instead of discrete labels (e.g. 1-100 /bAk/ or /dAk/), the distribution of ratings was more consistent with a model of continuous perception (Massaro & Cohen, 1983). In fact, the effect of the instruction manipulation mentioned earlier, in which prompting participants to perceive the same stimuli as either speech or whistle sounds has been in the empirical record for a long time, at least since the early 1960s (reports of this effect can be found in Lane, 1965; Mattingly, Liberman, Syrdal, & Halwes, 1971; Pisoni & Lazarus, 1974; M. E. H. Schouten, 1980). In a more recent paper B. Schouten, Gerrits and van Hessen (2003) argue for "The end of categorical perception as we know it". Their data show that categorical perception of speech sounds depends on specifics of the discrimination task used (see also Gerrits & Schouten, 2004). Many factors like inter-stimulus interval, "naturalness" of the stimuli and task induced response bias were found to determine whether categorical perception and phoneme boundaries were observed.

The origin of these divergent results and their interpretations are due to the following apparent contradiction: We know the universe we perceive is continuous, but our perception of the universe appears to be discrete and not a continuum of "blooming, buzzing confusion" as William James (1890) called it. We perceive objects, people, words, and colours. It is no surprise then that the issue of whether perception is discrete or continuous, categorical or not, is by no means resolved since James argued in favour of continuous perception (see Spivey & Dale, 2006 who argue for the continuity of mind, whereas VanRullen & Koch, 2003 suggest discreteness of perception is a neurological reality). The idea that speech exists in our nervous system as static segmental building blocks that represent collections of discrete phonetic features that serve perception as well as articulatory gestures by sequential alignment, is an attempt to shape a scientific theory about the "linguistic world" in the way certain humans who are literate in alphabetic writing systems experience that linguistic world (Port & Leary, 2005; Port, 2010a). This is an example of the conjectures in Chapter 1 in which I suggested that theoretical constructs are often confused for constituents of reality. Lisker and Abramson gave an excellent description of this biased view of reality some 40 years ago when discussing the Chomsky and Halle –linguist– theory of speech:

> "Their concern is not how an articulatory sequence and its associated acoustic signal, both of them physically neither purely continuous nor purely digital in nature, are related to a linguistic expression, but rather to impose digitalization on the physical description in such a way that it will necessarily be a description of the segments in the linguist's spelling of the expression." (Lisker & Abramson, 1971, p. 781).

5.3.2 Reprise: Causation

The developmental perspective, how speech production and perception emerge in the developing child is usually conceived of as suppressing certain predispositions and attuning others in order to achieve adult performance. What these predispositions are varies; some authors suggest that innate feature detectors for complex phonetic segments are activated and deactivated during development (Werker & Tees, 1984). Others argue that phoneme categories emerge as perceptual magnet effects caused by innate speech sound prototypes (Iverson & Kuhl, 2000; Kuhl, 1991). The fact that the examples given so far reveal strong context sensitivity undermines the 'nativist' explanations that have been put forward. Why imagine biological and neurological structures into existence that turn out to be used only under some special experimental conditions? The logic from the point of view of a theory may be understandable: Elementary particles, the phonemes most literate adults intuit to exist, must be constructed from smaller elementary particles in the pre-linguistic child. That way the object of development is to construct the larger particles from the smaller ones, and the differences between infants and adults are explained:

> "What the child will need to acquire with experience is the appropriate criterion for classification (i.e., the appropriate boundary value) and the ability to ignore, at least under some circumstances, acoustic differences that signal variants of the same phonemic category" (Eimas & Miller, 1980). "Choke on that causality!" (Farnsworth, 2001).

To summarise, there are problems with theories maintaining that speech sounds are represented as discrete components of invariant static acoustic features. These problems arise because the speech signal is predominantly a continuous physical signal, because the complex gestures that generate the signal appear to be under continuous control by the speaker and because it is likely that the segments appear in scientific theories due to scientists being literate in alphabetic writing systems. Although perception along a physical continuum may appear to be categorical in most cases, perception of discrete category boundaries is highly context dependent and under experimental control. A listener, depending on subtle contextual variations, may report of the same physical features of a signal that they cause different sensory experiences. Moreover, there are clearly substantial individual differences in perceived and produced phoneme boundaries by native speakers of the same language, which contests the assumption that static possibly innate phoneme boundaries exist.

5.4 Critical Phoneme Boundaries vs. Coordination Dynamics

There are several contemporary theories and models that provide a dynamical, continuous and interaction-dominant account of speech perception and production that deal with some of the problems with component based theories mentioned above (van Lieshout, 2004 is a review). In the present study I will expand an existing nonlinear dynamical model of speech perception (Case, Tuller, Ding, & Kelso, 1995; Tuller, Case, Ding, & Kelso, 1994; Tuller, 2004) to account for the emergence of perceptual boundaries and categories along a physical continuum without the need to invoke abstract segmental components to represent each perceived perceptual category.[1] Another motivation to consider this model was to find an alternative to the allophone coupling hypothesis of the pre-verbal to verbal to literate development argued for by Serniclaes and colleagues in order to explain better within phoneme category discrimination by dyslexic readers.

Tuller et al. (1994) have shown that critical phoneme boundaries in the perception of a VOT contrast are actually rare. In a classical speech categorisation experiment the stimuli that make up a synthetic continuum are presented in a random order and a category switch or phoneme boundary is usually found somewhere in the middle of the continuum. When the stimuli are presented in a sequential order up and down the continuum however (e.g., going gradually from /bAk/ [container] to /dAk/ [roof] and back) three phenomena are typically observed: Hysteresis, which is a delayed category switch, enhanced contrast, which is an early category switch (the reverse of hysteresis), or a critical boundary, which is a category switch at the same point along the continuum irrespective of the direction it was traversed. If speech perception were the mere detection of a collection of acoustic feature values, one would expect to always see a critical phoneme boundary. However, hysteresis and enhanced contrast are observed more frequently with sequential presentation of the stimuli (Case et al., 1995; Tuller et al., 1994).

To account for these phenomena Tuller et al. (1994) introduced a potential model that can explain the differences between individuals and contexts that decide when participants switch from hearing one speech sound category to another one. The mathematical form of the model is shown in equation 5.1 and 5.2.

$$\frac{dx}{dt} = -\frac{dV(x)}{dx} \tag{5.1}$$

$$V(x) = kx - \alpha\frac{x^2}{2} + \beta\frac{x^4}{4} \tag{5.2}$$

It is derived from a class of models used to understand the emergence of and switching between behavioural patterns or perceived categories known as the HKB-model of coordination dynamics (Haken, Kelso, & Bunz, 1985; Kelso & Schöner, 1988; Schöner, Haken, & Kelso, 1986).

[1] Many thanks to Ralf Cox who provided the mathematical description and initial MATLAB implementation of the model.

Recently the HKB-model has been implemented in the task dynamics framework (Saltzman & Byrd, 2000; Saltzman & Munhall, 1989) as part of a coupled oscillator model of speech gesture timing that was used to model the self-organization of syllable structures (Goldstein, Nam, Saltzman, & Chitoran, 2008; Nam, Goldstein, & Saltzman, 2009). This model is a so-called potential function describing which regions of a state space a dynamical system is attracted to given certain parameter settings. The minima of the potential function given in equation 5.2 represent the stable states of the differential equation given in 5.1. In other words, the potential model represents the potential outcomes of a dynamical continuous process as a discrete number of states. This is exactly the question that needs to be addressed by a model of speech perception according to Lisker and Abramson quoted above: How do dynamical, continuous signals relate to apparent static discrete outcomes of those processes? Another way to define the goal of a speech perception model is to understand how the graded internal structure of speech sound categories depend on the multitude of acoustical features generated by articulatory gestures (Case et al., 1995).

A potential model, when applied to speech perception, describes a dynamically changing landscape of attractor states, corresponding to a perceivable speech sound category (see Figure **5.1**). The number of available categories and their relative strength of attraction, or stability (reflected by the deepness of the wells) emerge dynamically during the task and are governed by a control parameter k. This is a collective variable that will be discussed in more detail below. Figure **5.1** shows what the potential landscape looks like for five different values of k. Suppose this is a model of the perception of a /bAk/ - /dAk/ continuum like the one discussed in Chapter 4. The two end points of the continuum are represented by the parameter values of $k = -1$ (/bAk/) and $k = 1$ (/dAk/). For these values there is one stable basin of attraction or one perceivable category. As k increases from -1 to 1 the stability of the perceived category /bAk/ (indicated by the ball in the potential well) decreases and a second trough emerges in the landscape. At some critical value of k the stability of the percept /bAk/ has decreased by such an amount that a category switch occurs as the ball settles into the more stable state representing the perception of /dAk/.

When $k = 0$ the potential wells associated with the two observable categories are both available, but very unstable. In this so-called bi-stable state many factors may influence which of the two attractors the system eventually settles into. One such factor is which category was perceived on the previous trial and it is this feature that Tuller et al. (1994) were able to confirm experimentally by showing that sequential traversal of the continuum results in hysteresis and enhanced contrast effects. Random presentation results in the observation of a critical boundary because, if randomisation was accurate, there is a 50% chance of having perceived either one of the categories on a previous trial.

The shape of the potential landscape does, however, not depend exclusively on the value of the acoustic parameter that is being manipulated to create the continuum. When exactly the system enters into this bi-stable state depends on the collective variable k. This variable depends on the number of previously perceived stimuli and certain participant characteristics. The equation for k is given by equation 5.3:

$$k(\lambda) = k_0 + \lambda + \frac{\varepsilon}{2} + \varepsilon\theta(n - n_c)(\lambda - \lambda_f) \tag{5.3}$$

In this equation, k_0 is the value of k for the stimulus at one end of the continuum, usually set at -1. In the case of a VOT continuum λ represents the voice onset times of a stimulus and λ_f would correspond to the VOT of the last stimulus on the continuum (the stimulus at the opposite end of the continuum from the one represented by k_0). For a standard VOT continuum λ is linearly equivalent to the acoustic feature that is manipulated, but this is not a prerequisite and nonlinear manipulations are in principle also possible (Case et al., 1995).

The parameter ε is described by Tuller et al. (1994) as summarising cognitive characteristics of the perceiver, like attention, linguistic experience, and learning. For the remainder of this Chapter,

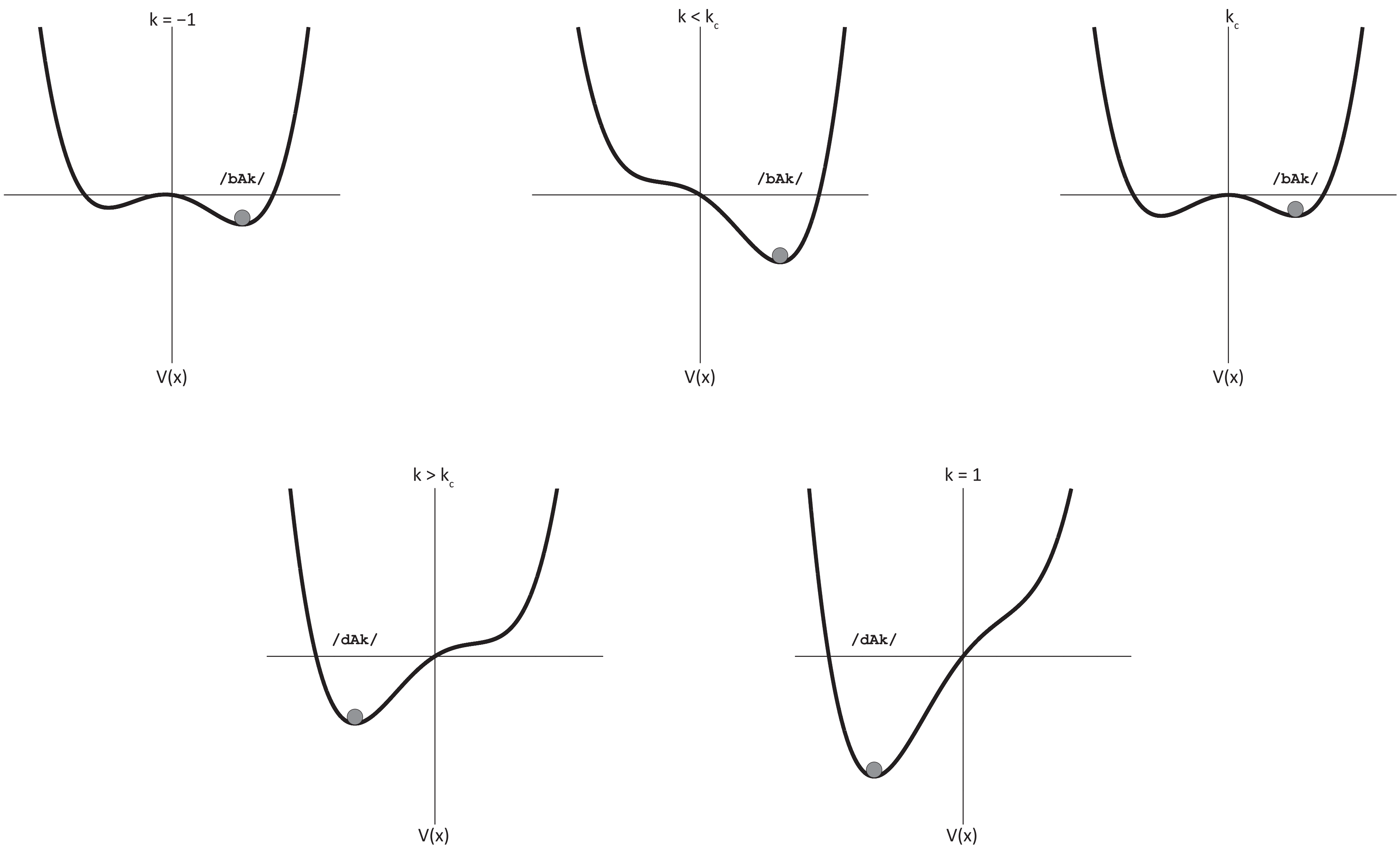


Figure 5.1 – The potential landscape of the model given in equation 5.2 for five values of k. The ball in the potential well represents which category is perceived. As k increases from -1 to 1 the relative stability of the perceivable categories changes.

I will interpret this parameter as explaining random or otherwise unexplained variation between individuals. The terms n and n_c are related to the number of stimulus presentations in the experiment: n just counts how many stimuli have been presented and n_c represent a critical number of stimulus presentations such that when $n - n_c \geq 0$ the step function θ changes from 0 to 1. In other words, when the number of presented stimuli reaches a certain number and $\theta = 1$, the value of k will be determined by an extra term $\varepsilon * (\lambda - \lambda)$ that regulates the impact of individual characteristics on k depending on how far the feature of the current stimulus is separated from its most extreme value along the continuum. This dependency of k on λ, ε and n results in an asymmetrical distribution of attractor stability once the number of presentations accumulates beyond n_c.

A hypothetical example of a sequential traversal of a continuum is shown in Figure **5.2**. Here n_c is set to 7, so when the continuum is traversed back from stimulus 7 to 1 the relative stability of percepts has increased. Stimulus 1 at the end of the run is more stable than stimulus 1 at the beginning of the run as indicated by the deeper potential well. This effect has been experimentally verified by Case et al. (1995), who asked subjects to rate to what extent a stimulus was representative of a category label. They found that a stimulus at the beginning of a sequence was rated less exemplary for a phoneme category compared to the exact same stimulus presented at the end of a sequence. Figure **5.2** also shows the effect of sequential presentation on the location of the category switch; going up the continuum from 1 to 7 the switch occurs between 5 and 6. Going back down the continuum the switch occurs between 5 and 4. There are several other interesting features of the model that may help understand the current empirical record of data related to speech perception by dyslexic readers. I will discuss them after I have introduced an extension of the model to two dimensions.

5.5 Interlude: Coupling of Higher Order Variables

A notion that is common in competing theories of language acquisition is consistent with contemporary interaction dominant theories of human development: An infant somehow has to couple information it acquires about the structure of the environment through different sensory modalities in order to act upon that environment (Bates, Thal, Finlay, & Clancy, 2002; Smith & Gasser, 2005). For speech perception and production a narrow view of this developmental task involves attributing meaning to the sounds caregivers use to communicate, as well as to acquire the necessary motor control to produce sounds that are similar to those of the caregivers. This is envisioned in the quote by Eimas and Miller cited above and the coupling hypothesis in the allophone perception theory of speech perception as mostly biologically determined. In dynamical theories of development, there is no developmental task, no blueprint or goal (e.g. Smith & Thelen, 2003). The infant is immersed in a world of multimodal sensory experience and through soft-assembled mechanisms gains knowledge about the structure of the environment (e.g. Kloos & Van Orden, 2009; Thelen, Kelso, & Fogel, 1987). Sensory systems overlap and are time-locked (correlated sensations) which is how they inform each other about the structure of the physical world without a supervisor, or blueprint. These are the concepts of degeneracy and re-entry associated with neural networks that receive multimodal input (Edelman, 1987; Tononi, Edelman, & Sporns, 1998). Degeneracy refers to the fact that any function can be carried out by more than one configuration of a neural system and that the same clusters of neurons participate in different functions (i.e., in contrast with static components that serve one purpose). Re-entry refers to the fact that activation of sensory experience by one modality can activate other modalities through the time-locked coupling between sensory and action systems (Smith, 2005).

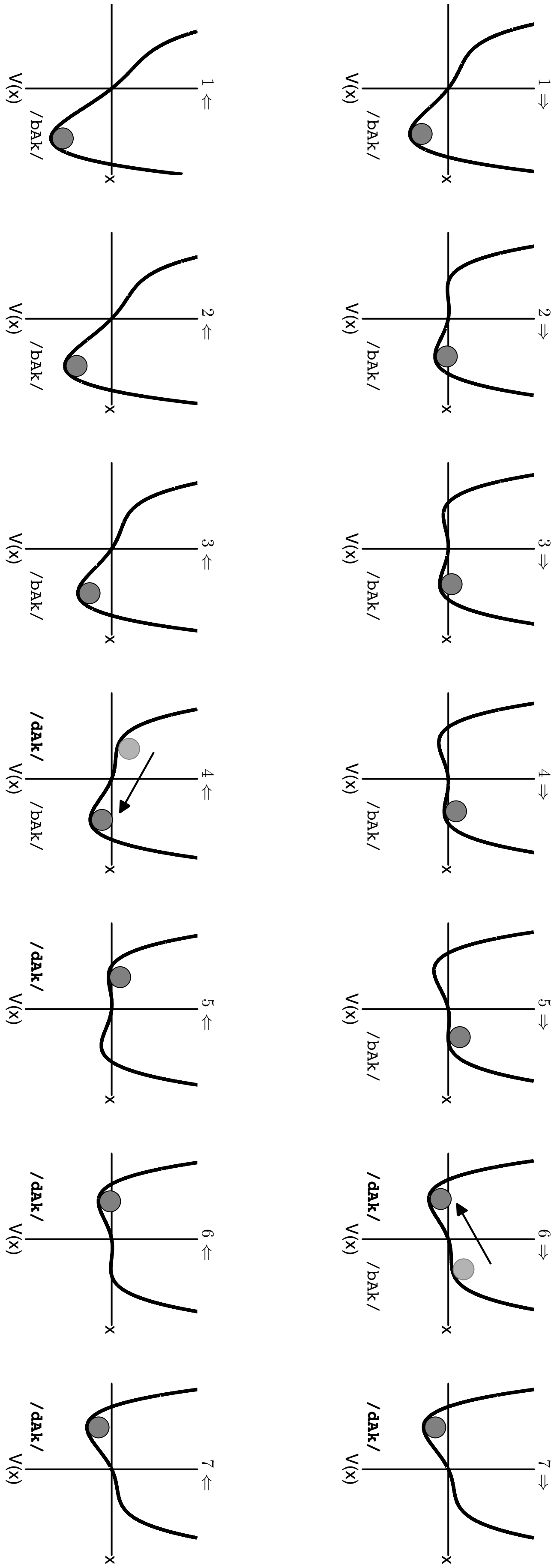


Figure 5.2 – A hypothetical sequential run across a continuum from /bAk/ to /dAk/ and back to /bAk/. The effect of a delayed category switch can be seen, as well as the increased stability of the perceptual category for presented stimuli at the end of the sequence. See text for details.

5.5.1 Multimodal Language and Literacy Acquisition

Speech perception is multimodal from the very beginning, but involves not just auditory perception and speech gestures (e.g. Bates & Dick, 2002). Imagine an infants' early sensory experiences related to speech: Whether or not eye contact is established, many adults will spontaneously start to utter sequences of syllable-like sounds (or other forms of motherese), accompanied by exaggerated pitch and prosody and gestures of the lips. Often the rhythm of the sequence of utterances is accentuated by tapping or by rocking the infant. Self-initiated rhythmic movements to sounds aid infants' perception of beat (Phillips-Silver & Trainor, 2005) and several lines of research suggest perception of rhythm may actually underlie speech segmentation (Nazzi, 2003; Ramus, Hauser, Miller, Morris, & Mehler, 2000; Ramus, Nespor, et al., 2000; Thomson & Goswami, 2008; Tincoff et al., 2005). Three months after birth infants start to babble and after 7 months they can imitate sounds with consonants (Bates et al., 2002). This self-assembled 'tuning' to the environment (adaptive behaviour) has been modelled in the DIVA model of speech production. It mimics development of native speech sound production by coupling information from different modalities: A babbling mode that produces random speech-like sounds which can be 'tuned' by feedback from the auditory input to the model, for example a 'caregiver' repeating a native sound over and over (Callan, Kent, Guenther, & Vorperian, 2000; Guenther & Perkell, 2004).

After 8 months infants reportedly start to loose the ability to discriminate between non-native speech contrasts, a process of which the allophonic mode of speech perception hypothesis suggests it does not unfold as it is supposed to, due to a failed coupling of allophones to native phoneme contrasts in dyslexic readers (Serniclaes et al., 2004; Serniclaes, 2006). This is, however, not a description of a multimodal coupling and appears to be purely a coupling of auditory categories, mental representations. What is missing from the hypothesis is the role of speech production. Infants will establish perception-action couplings when they start to babble and produce consonants. In most theories of development these perception-action couplings are thought to be the driving force behind achieving developmental milestones and they may involve coupling between basic sensory experiences or concern more complex, higher-order variables (Adolph, 1997, 2008; Camaioni, Aureli, Bellagamba, & Fogel, 2003; Hsu, Fogel, & Cooper, 2000; Thelen, 1996, 2000; von Hofsten, 2004).

5.5.2 Multimodal Language and Literacy Production

Ever since the first experimental results were published that showed that speech perception may be related to speech production (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Liberman, Delattre, & Cooper, 1952), the motor theory of speech perception has gone through several revisions. An extensive recent review concludes that the empirical record shows strong evidence that perceiving speech is perceiving speech gestures involving the motor system (Galantucci, Fowler, & Turvey, 2006). Of course, infants have already been learning about the structure of rhythmic sounds and speech sounds in the womb (DeCasper & Spence, 1986; Smotherman & Robinson, 1988), some time before being able to produce sounds. Given the radical view on phoneme representations as perceptual ghosts of letters (Port, 2008a, 2008b), by analogy one might argue that perceiving phonemes is perceiving letters and involves an alphabetic writing system. It would be odd to have the writing system internalised as a collection of representations, letters are visual structures in the environment, and sounds are audible structures in the environment. Reframed in this way pre-verbal to verbal development and pre-literate to literate development both require coupling between structural regularities of the world accumulated in different biophysical subsystems and these should be bi-directionally coupled. Correlates of speech gestures are indeed found in studies of reading. There are for instance effects of the way a word is pronounced on the way a word is read in terms of speed, accuracy or semantic categorisation (Bosman & De Groot, 1995; Bosman &

de Groot, 1996; Bosman, van Leerdam, & de Gelder, 2000; Dijkstra, Grainger, & van Heuven, 1999; Van Orden, 1987).

To summarise, speech gestures are embodied and therefore by definition in some way biologically coupled to perception within an individual. Acquisition of speech sound production appears to occur relatively unsupervised, or at least there are computational models using feedback mechanisms that give a plausible account of how such a bi-directional coupling may give rise to production of native speech sounds. At first glance learning to read an alphabetic writing system seems more like acquiring an abstract set of correspondence rules, something artificial, cultural and disembodied. Learning to read, however, is as multimodal and action driven as learning to speak; it involves perception of visual regularities as (sequences of) letters, coordination of eye movements to speech gestures when reading aloud and fine motor control for speech symbol production (writing). Somehow 'learning grapheme to phoneme correspondence rules' does not quite capture the full complexity of the developmental process of becoming literate.

5.6 A Multi Dimensional Potential Landscape

The notion that coupling of dynamical systems gives rise to new joint patterns and behaviours is a well known and well studied phenomenon in physics and mathematics, but has also been applied in neuroscience and behavioural science as synchronisation and coordination of physiological processes and behaviour (e.g. Glass, 2001; Nam et al., 2009; Nowak & Vallacher, 2003; Schöner & Kelso, 1988; Shockley, Baker, Richardson, & Fowler, 2007). The key property that is interesting for the purpose of understanding speech perception by dyslexic readers is that coupled dynamics constrain the degrees of freedom of the coupled systems, depending on the strength and nature of the coupling. This also means that less coupling means more degrees of freedom, or more stable states the system might settle in.

Consider a coupling of two potential models that represent higher-order perceptual variables; one describes the observed auditory categorical perception of stimuli on a certain physical dimension (e.g. Jusczyk et al., 1977), whereas the other describes a similar continuum, but by self-produced speech gestures. It has been shown that transitions between stable articulatory gestures can be described by coordination dynamics in terms of self-organisation, control parameters, critical values, and bi-stable states (Kelso, 1986; Tuller & Kelso, 1991). In fact, the value of the control parameter at which an individual switches from producing one stable articulatory state to another (e.g., pronouncing /ip/ at increasing rates until it becomes /pi/), occurs at the same control parameter value as a switch in the perceptual state indicating strong coupling between the perceptual and motor systems (Kelso, 1995; Munhall et al., 1994). More recently, a formal notion of multimodal coordination dynamics has been put forward (Lagarde & Kelso, 2006). For the purpose of this chapter it is sufficient to consider coupling between two higher order variables as discussed above. The assumption is that the coupling of these two models of coordination dynamics are indicative of the quality of integration between speech perception and production of a particular physical description of the speech signal as evidenced by Perkell et al. (2004).

The result is a two-dimensional potential landscape whose shape is now also governed by the way the two functions are coupled. The set of two ordinary differential equations that describe this system is:

$$\frac{dx}{dt} = -\frac{\partial V}{\partial x} \quad \frac{dy}{dt} = -\frac{\partial V}{\partial y} \tag{5.4}$$

Equation 5.5 is the two-dimensional potential model written in (multivariate) polynomial form, which may be more familiar to the behavioural scientist.

$$V(x,y) = a_{10}x + a_{01}y + a_{11}xy + a_{20}x^2 + a_{02}y^2 + a_{40}x^4 + a_{04}y^4 \tag{5.5}$$

The model as used in this study is given in equation 5.6. Several terms have been rearranged and some of the parameters that remain constant have been filled in.

$$V(x,y) = k(x-y) + \gamma_{cs}(x*y) - \frac{(x^2+y^2)}{2} + \frac{x^4+y^4)}{10} \tag{5.6}$$

The parameter k has the same function as in the original model; it is a collective variable that represents the value of an acoustic parameter of a speech sound combined with a weighted average of previously perceived stimuli and characteristics of the individual. It is in principle possible to have one k parameter for each dimension. A high coupling strength (γ_{cs}) means the joint dynamics of the two models will yield two major stable states or categories that will follow roughly the same dynamics under the influence of k as in the one dimensional potential model. The potential landscape for $k = 0$ (the bi-stable state) is shown for different values of γ_{cs} in Figure **5.3**. It represents a coupling hypothesis for speech perception in developmental dyslexia that does not presume innate detectors of complex stimuli. It does have the potential (pun intended) to explain the same phenomena the allophonic mode of speech perception hypothesis seeks to explain.

When there is no coupling between the perceptual system that perceives a physical dimension of a complex speech signal and the action system that can produce the complex signal, as is the case in the prelinguistic or preverbal child, there are four basins of attraction for $k = 0$. The bi-stable state is thus rather a multi-stable state in this context. Suppose a preverbal infant was examined for the discrimination of stimuli with different values for this particular physical dimension, it would very likely result in three -not four- different switches between stable percepts. This can be explained by examining Figure **5.4**, which shows a decoupled model for different values of k. The middle row shows a projection of the line $x = y$, a cross-section through the potential landscape. The bottom row shows a contour plot of the potential landscape, the arrows follow the local decreasing potential gradient. Imagine again a ball representing the percept the system settled into, starting at /bAk/ with $k = -1$ (left column of Figure **5.4**).

Under influence of the changing control parameter k the multi-stable state is approached ($k = 0$) and eventually the critical value of k where a category switch will occur. This is shown in the fourth column ($k > k_c$). There are three possible gradient paths to follow leading to a new stable state. The ball could follow the line x=y and end up in the deepest well (/dAk/). Figure **5.4** shows a situation where the system settles in a stable state in the upper right corner of the contour plot. As k increases, another category switch will occur when the state becomes too unstable and just one stable state remains ($k = 1$). For high coupling strengths the relative strength of the two major stable states exceed the less stable states by such an amount that the probability of the system settling in in one of them is very low. This probability is not zero however, which would correspond to data that point towards very weak within category discrimination by adults interpreted as evidence for "natural" innate allophone boundaries (Serniclaes, 2006).

Based on this model the hypothesis is that a coupling of perception and action systems decreases the degrees of freedom available to the joint system to discriminate between values along the physical continuum. This explains how in uncoupled, or loosely coupled systems, more category boundaries may be observed. These categories are not "things", components that are stored somewhere, nor do they represent static values of the physical dimension. They emerge as more or less stable states from the interaction dynamics. When and where category switches are observed depends on many contextual factors, such as how many stimuli have been perceived, cognitive characteristics of the individual and the strength and nature of the coupling. The nature of the coupling as implemented here is just one of many possible ways to achieve joint dynamics. In the present model it results in a symmetrical, orthogonal 2D landscape.

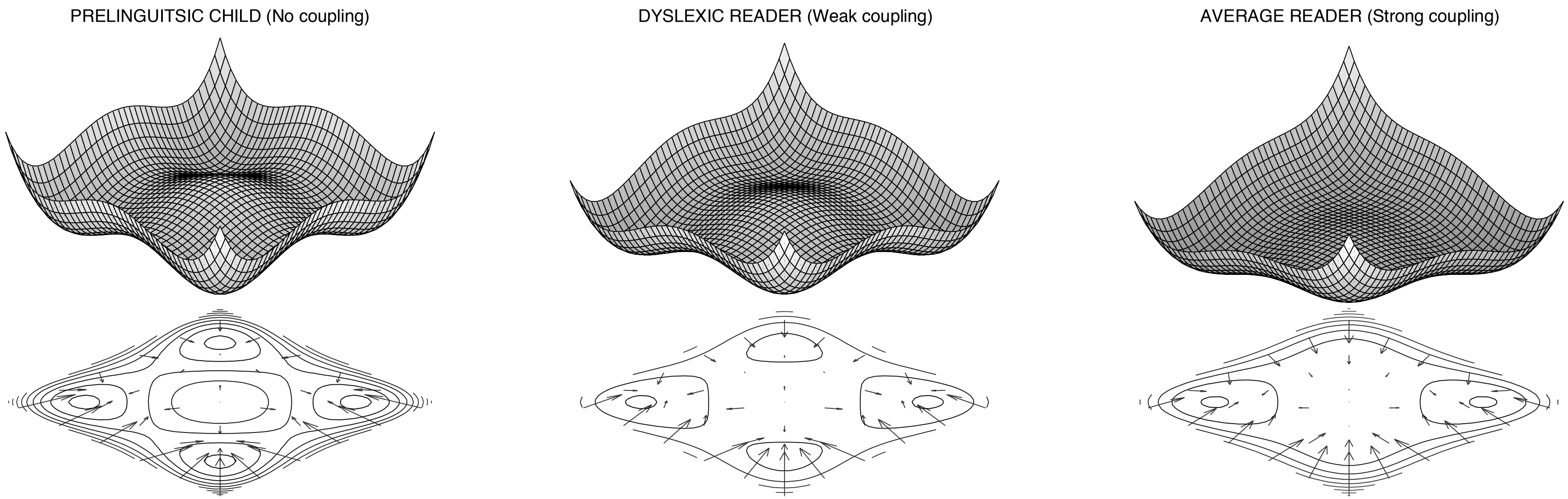


Figure 5.3 – The figures show how different coupling strengths between collective variables shape a 2D potential landscape for $k = 0$. There are more and less stable basins of attraction projected as contour maps below the landscape. Here the potential wells correspond to stable percepts, interpreted as elementary speech particles like allophones or phonemes by component theories of speech perception.

The uncoupled two-dimensional potential model for different values of k (columns). The top row shows the potential surface. The middle row shows the projection of the potential surface along the line $x = y$. The bottom row shows contour and quiver plots. The orientation of the arrows may be interpreted as the direction a ball would roll towards were it dropped into this landscape. The length of the arrows may be taken as representative of the velocity with which the ball would depart in that direction.

5.7 Hypotheses: Testing Model Predictions for Dyslexic Readers

Is it reasonable to assume that dyslexic readers may have a problem coupling motor and sensory systems? There is a large body of work, including neuroscientific evidence, suggesting a cerebellar deficit that might play a role in developmental dyslexia (Nicolson & Fawcett, 2006). This deficit is thought to affect motor skills (Nicolson, Fawcett, & Dean, 2001; Thomson & Goswami, 2008), procedural learning (Nicolson, Fawcett, Brookes, & Needle, 2010) and balance (Brookes, Tinkler, Nicolson, & Fawcett, 2010; Stoodley, Fawcett, Nicolson, & Stein, 2005). For now let us interpret these results as supporting the plausibility of the coupling hypothesis for developmental dyslexia, given the contemporary view on the role of the cerebellum in coordination and control of movement and its densely connected sensory-motor circuit (Balsters et al., 2010; Gowen & Miall, 2007). More particularly, because of the recent finding that reveals the important role of the cerebellum plays in speech perception and production, such as the planning and temporal organisation of speech gestures (Ackermann, Mathiak, & Riecker, 2007). Based on the potential model at least two hypotheses can be derived that predict effects on the speech perception of dyslexic readers that would be very difficult to understand from the perspective of theories based on innate feature detectors, or feature values like the allophonic mode of perception hypothesis.

The first hypothesis is that the better within-category discrimination ability of dyslexic readers, interpreted as perception of a rudimentary allophone boundary, may also be caused by perturbations of the system other than evaluating whether the acoustic features of a stimulus are smaller or larger than values that make up this innate boundary. Because one perceptual category looses stability and others gain stability, adding ‘noise’ to the system might cause it to organise into a new stable state. One way Tuller et al. (1994) established such a noise induced category switch was by sequentially approaching the multi-stable state and repeatedly presenting the same stimulus at a point along the continuum where a category switch might be expected. They found that participants were inclined to report perception of the other category after 2-6 repetitions of the exact same stimulus, representing one and the same value of the physical dimension under study. When this effect is translated to discrimination of values of the continuum as ‘same’ or ‘different’ stimuli, one might expect an additional discrimination peak in a system with more than two stable states (as explained in Figure **5.4**) in noisy conditions.

In previous chapters it was shown that the acoustic manipulations of stimuli, especially the amplification of fast changing spectral components and both slowing down and amplification of the stimuli acted as a reduction of discriminability of stimuli and an increased bias of perception of one category over the other. The first hypothesis will be tested by an experiment in which stimulus pairs from the same /bAk/ to /dAk/ continuum as discussed in Chapter 4 will be discriminated by average and dyslexic readers using the four conditions of acoustic manipulation: None, Slowed Down, Amplified and Both. The prediction is that the dyslexic readers will show within category discrimination peaks not observed in average readers. The discrimination peaks are expected for acoustic manipulations that have previously been shown to disrupt, or destabilise perception: Amplification and Both.

The second hypothesis concerns the effect of coupling strength on the internal structure of the perceived categories. First, it may be deduced from the multimodal coupling hypothesis that cou-

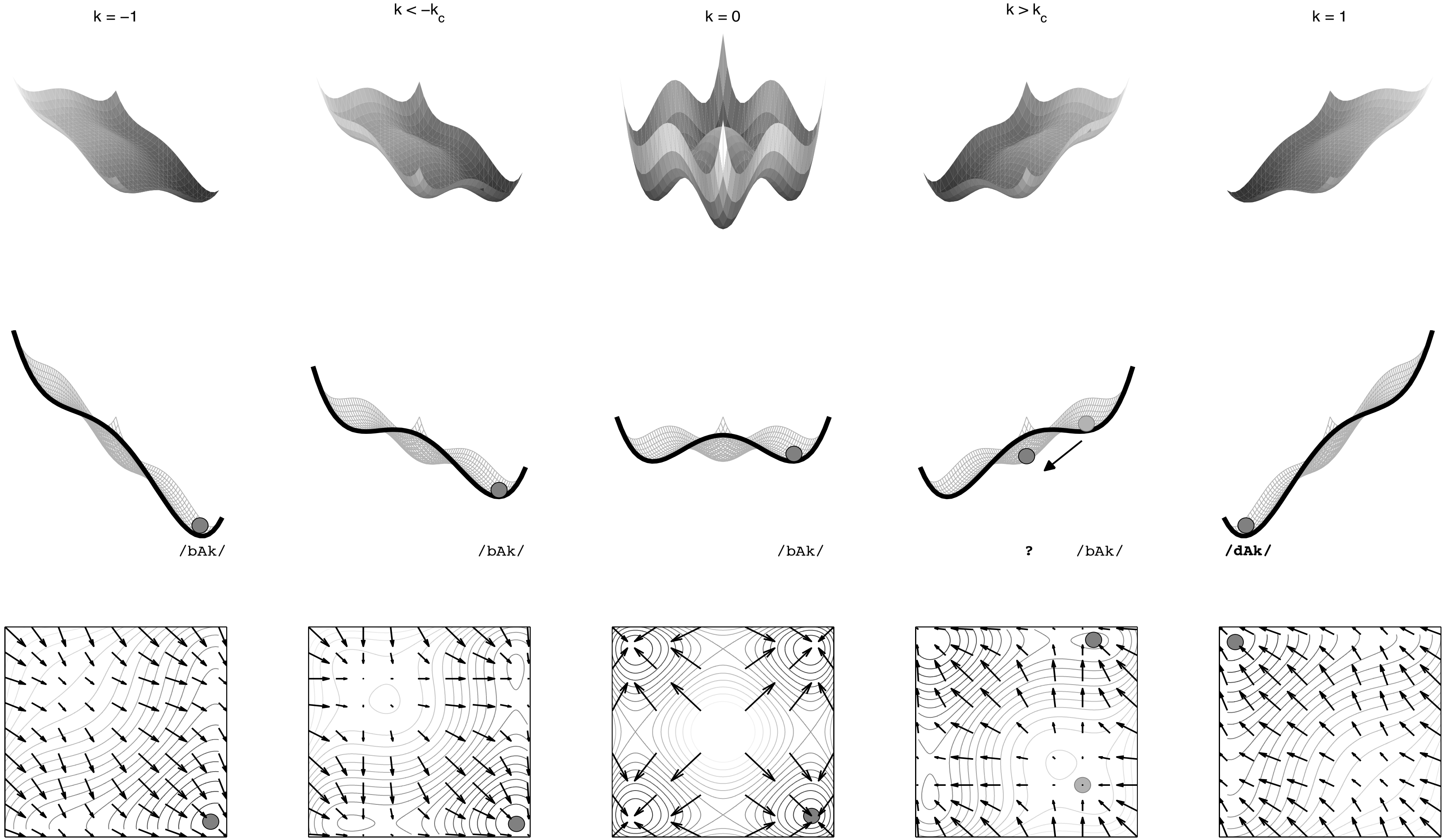


Figure 5.4 – The figures show how different coupling strengths between collective variables shape a 2D potential landscape for $k = 0$. There are more and less stable basins of attraction projected as contour maps below the landscape. Here the potential wells correspond to stable percepts, interpreted as elementary speech particles like allophones or phonemes by component theories of speech perception.

pling strength will increase during development due to gained linguistic experience, including literacy training. This suggests the within-category discrimination ability by dyslexic readers should decrease with age and experience and this is in fact not contradictory to the allophone coupling hypothesis (Bogliotti, Serniclaes, Messaoud-Galusi, & Sprenger-Charolles, 2008; Serniclaes, 2006). Predictions based on the potential model are more precise and concern a decrease in the occurrence and/or size of hysteresis and enhanced contrast due to an increase in the stability of two states at the cost of less stable states. In other words, the internal structure of the categories will become more homogeneous as the coupling strength increases. This trend should be observable in both dyslexic and average readers. The change in internal structure will be examined by looking at the distribution of transitions for three different age groups. Dyslexic and average readers labelled a 20 step continuum in three conditions: (i) Sequential /bAk/ » /dAk/ » /bAk/; (ii) Sequential /dAk/ » /bAk/ » /dAk/ and (iii) Random. The predictions are that within the dyslexic and average reader groups, homogeneity of the distribution will increase with age as suggested by the coupling hypothesis. Between those groups, the youngest children are expected to show the clearest differences in transition distributions between dyslexic and average readers. The distributions of older dyslexic readers should, however, resemble the distribution of the younger average readers.

The third and final hypothesis concerns a test of the 2D potential model as a tool for statistical inference. Can model parameters be inferred from observed data and will the fitted values be in accordance with the hypotheses and predictions formulated so far? The method used to fit the data will be described and will concern of course the coupling strength (γ_{cs}), but also the parameters that are associated with individual differences and may vary within participants during the course of an experiment. These parameters are the critical number of stimuli (n_c) that governs when the step function changes from 0 to 1. This causes an asymmetric slope of the potential across the entire sequence of presented stimuli such that stimuli presented at the end of the sequence will have deeper potential wells than those presented at the beginning of the sequence.

How large this effect will be, which basically involves the size of hysteresis and enhanced contrast jumps, is based on individual differences and is captured by the parameter ε. This suggests that the location of any perceptual boundary, allophonic or otherwise, is expected to change within a participant, most notably under influence of the way in which the continuum is traversed (sequentially or in a random order). Moreover, in a loosely coupled system there should be more evidence of unstable categories than in a strongly coupled system. And finally, coupling strength estimated from the data should be related to age and reading ability analogous to Experiment 2. In order to test this, the data from Experiment 2 will be used to estimate the parameters for each trial sequence separately. Subsequent analyses on the parameter estimates must reveal whether the parameter estimates are in accordance with the predictions so far.

5.8 Test 1: Allophonic boundary perception or noise perturbation?

5.8.1 Method

Participants

The participants[2] in this study were the same as described in Chapter 4. The data were collected in a two sessions on the same day. There were 78 children with valid data[3] (age range 101.2 to 159.3 months) from 9 different schools in the southeast of the Netherlands. Half of the participants (39)

[2] I would like to thank the students who participated in the Research Seminar 2004 of the School of Pedagogical and Educational Science for their aid in data collection.

[3] One participant who completed both the identification and discrimination studies was removed from the present analysis (ID 804 in the data file, available here https://osf.io/a8g32). The data that was recorded in the output file for the speech discrimination experiment was partially missing.

were dyslexic readers as indicated by two reading tests: A timed-reading task for regular words ("Drie-Minuten-Toets"; Verhoeven, 1995) and a timed pseudo-word reading task ("KLEPEL"; van den Bos et al., 1994). When the child's scores on both tests were within the 25th percentile (norm score by age), the child was considered to have severe reading problems (see Table **4.5**).

Stimuli and Acoustic Manipulations

The stimuli were based upon natural speech recordings for the words /bAk/ [container] and /dAk/ [roof] and transformed to create a 10-step /bAk/ to /dAk/ continuum (Van Beinum, Schwippert, Been, Van Leeuwen, & Kuijpers, 2005) using the Praat program (Boersma & Weenink, 2005). The stimuli differed only with respect to the second formant transition of which the onset frequency is gradually increased from /bAk/ to /dAk/ (see Table ref for exact values)

All the stimuli of this F2 continuum were manipulated in three manners using the Praat program (Boersma & Weenink, 2005). First, the speech signal was Slowed Down to 150% of its original length. This was achieved by a Pitch Synchronous Overlap and Add (PSOLA) algorithm (see e.g., Segers & Verhoeven, 2005). Second, the signal was Amplified by 20dB, but only the fast changing spectral elements. The algorithm used to do this in Praat was similar to the one used by (Nagarajan et al., 1998), who confirmed this in a personal communication with Segers and Verhoeven (2002). Third, Both manipulations were applied as is done in the FastForWord program (Nagarajan et al., 1998; Tallal, 2004): The speech signal was slowed to 150% of its original length and all the fast transitional elements were then amplified by 20dB. There was of course also a continuum that had None of the manipulations applied to it. This yielded 40 different stimuli in total.

5.8.2 Procedure

Speech perception experiment

The speech discrimination task was presented on a laptop computer in a quiet room at the children's school. In the discrimination task, two smiley faces appeared on the laptop screen. First one face uttered a word and then, after 500ms, the second face uttered a similar or dissimilar word. The children were told that the second face was trying to say the same word as the first face had said. They were then asked to listen very carefully and decide whether the second word was the same or not. After the utterance of the two words by the smiley faces, two frames with pictures in them appeared on the left and the right of the screen. One of the frames contained a picture of two green smiling faces. The other frame contained a picture with a green smiling face and a red frowning face. The children were told to press a designated key on the keyboard corresponding to the left or right frame: When the two words were the same they had to press the button corresponding to the two green smiling faces and when the two words were not the same the button with the mismatched faces. The pictures in the frames were randomly interchanged upon each presentation. Prior to the experimental trials, 10 practice trials were presented in which the participants received feedback. During the experimental trials, the three types of manipulated /bAk/ and /dAk/ stimuli and the unmanipulated stimuli were presented in same and different pairs (pairs were always separated three steps n the continuum). The order of the stimuli of the different pairs was randomly varied. Each stimulus pair was presented a total of 2 times resulting in 100 stimulus presentations (2 x 4 manipulations x 10 same/different pairs).

Statistical Analysis

For each participant there were 80 responses of either same or different. These data were entered in a logistic multilevel model (using MLwiN version 2.1; Rasbash, Charlton, Browne, Healy, &

Box 5.1: Measures extracted from a confusion matrix. They are used to construct the ROC and PR curves.

		Expected	
		different	same
Observed	different	TP	FP
	same	FN	TN

True Positive Rate (Sensitivity, Recall) = $\frac{TP}{TP+FN}$
True Negative Rate (Specificity) = $\frac{TN}{TN+FP}$
Positive Predictive Value (Precision) = $\frac{TP}{TP+FP}$

Cameron, 2009) with the 80 measurement occasions representing responses to a random permutation of the pairs on the ordered F2 continuum at level1. The stimuli were coded in such a way that all the same-pairs were 0, the first different pair (stimulus 1-3) was 1 and the last different pair was 7 (stimulus 7-10). This means the model is evaluated relative to the same-pair responses located at the intercept. The responses at the level of the measurement occasions were considered binomially distributed as 0 and 1 and a logit link function was used. The repeated measurements can be thought of as clustered within the participant, who represent a second level of random variation in the model (level2). The modelling strategy was as follows: First it was examined whether the multilevel model gave a better fit than a single level model with just measurement occasion defined as a level. Then, the empty multilevel model for change was fitted (M0), which in the present case means that the zero inflated fixed effect predictor as described above was added representing the rank order of the stimulus pair on the continuum (0-7). In a subsequent model (M2) it was examined whether stimulus rank could explain random variation in the slopes of the curve at the level of the participants (level2). If so, this means that the variation in discriminating between stimulus pairs between participants, can be understood as random variation with respect to the average discrimination curve of the entire sample. In the next step (M3) level1 and level2 covariates were added: a dummy variable that represents the four manipulation types (level1), and a dummy variable that represents whether participants are dyslexic or average readers (level2). In the final modelling step (M4) various interactions were tested including cross-level interactions between participant type and (Browne, 2009) stimulus type. The models were fitted using Monte Carlo Markov Chain simulation with 150,000 iterations. This number was chosen after inspecting the Raferty-Lewis diagnostic for each parameter estimate at each modelling step and was found to yield a very safe margin for all predicted parameters.

Classifier performance: Threshold averaged ROC analysis

The predictions from the multilevel model will be analysed in the context of classifier performance by generating Receiver Operator Characteristic (ROC) curves and Precision-Recall (PR) curves (Davis & Goadrich, 2006; T. Fawcett, 2004 are tutorials). These analyses are based on a confusion matrix containing hits and misses of a classifier detecting the "true" or expected class of a stimulus (binary decision problem). In the context of the present experiment the classifier types are the two groups of participants (average and dyslexic readers) who need to detect whether a stimulus pair is "same" or "different" for several values of the difference in acoustic parameter. The predicted responses of the multilevel logistic model for each participant to each stimulus are considered different thresholds of classifier performance. Their averaged classification curves (hence: threshold averaged ROC analysis) will be used to investigate whether some stimulus type x classifier type combinations show evidence of within category discrimination. The sample averaged observed responses to each stimulus pair serve as the targets in the analysis.

Box **5.1** displays a confusion matrix that can be constructed from the predicted responses of the multilevel model for each participant and "different" (or in ROC terminology: positive) stimulus pair by calculating true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) (Tripepi, Jager, Dekker, & Zoccali, 2009 contains an example of the use of logistic model predictions as classifier outcomes in ROC analysis). Each participant or threshold generates a series of coordinates in ROC space for each stimulus pair. These coordinates represent the True Positive Rate (also known as Sensitivity or Recall) and the False Positive Rate (1-Specificity). The interpolated curve through these coordinates is the ROC curve for a particular threshold. Imagine for example a medical diagnostic procedure that uses results from a blood protein test to make a decision about the presence of a certain medical condition. It might require Sensitivity greater than 70% (TPR > .7) and Specificity greater than 30% (FPR > .7) to detect the medical condition. ROC curves could help decide for which thresholds of the blood protein test results the diagnostic procedure achieves the required sensitivity and specificity (Duca, 2005; Tripepi et al., 2009).

The Precision-Recall space is different from ROC space in that the True Positive Rate (now dubbed Recall) is spanned against the Positive Predictive Value (also known as Precision). In many cases, curves in PR space provide extra information about the classifier performance when dealing with highly skewed distributions of positive and negative responses (Davis & Goadrich, 2006). This is likely to be the case in the present study where many of the pairs that physically differ from each other are expected to be perceived as “same” due to the categorical nature of speech perception. Based on studies in Dutch using the same (unmanipulated) stimuli as in the present study it may be expected that just one or two stimulus pairs will be judged as different on the level of the sample (Van Beinum et al., 2005).

Several statistics can be obtained from the curves, the most important one being the Area Under the Curve (AUC). A higher AUC means a better classifier performance. Moreover, each stimulus pair may be represented by a coordinate pair on the threshold averaged curve in ROC or PR space (Fawcett, 2004). This way the curves also allow direct comparisons of stimuli with respect to their location in ROC or PR space. The hypothesis is that the different acoustic manipulations will yield different curves, possibly in interaction with experimental group. Moreover, it is expected that the Amplified or Both stimuli will trigger a change in discriminability of some stimulus pairs, not observed with None and Slowed Down stimuli, especially in the group of dyslexic readers. This may manifest in discontinuities in the ROC of PR curves, or in the relative order of the stimulus pairs on the curves. In order to make comparisons between stimuli and experimental groups the statistics are bootstrapped to get confidence intervals around the estimates (based on 1500 resamples).

5.8.3 Results

Multilevel Logistic Model & Threshold Averaged ROC Analysis

The multilevel modelling strategy converged to the final logistic model M4 described above under Statistical Analysis. The model contains all the main fixed effects and their interactions: Pair(7) x Type(4) x Group(2). Due to the large number of fixed effects in the model Table **5.1** just shows the values of the Bayesian Deviance Criterion (DIC) for models M0-M4 described above (the steps are the same as in Table tab:6, but the rows reporting effects are more numerous). The predictions from M4 and the results from threshold averaged ROC analyses are shown in Figure **5.5** and Figure **5.6** . The line graphs represent the predicted probabilities for perceiving a stimulus pair as “different”. The 95% confidence intervals between average and dyslexic readers overlap for all stimuli in conditions None and Slowed Down (Figure **5.5**). There are no significant ROC and PR space differences in the areas under the curve (Table **5.1**). The ROC and PR plots in Figure **5.5** show that pair 6 (None) and pairs 1,6 and 7 (Slowed Down) in general trigger most unbiased responses from the average and dyslexic readers (as they are closest to the diagonal line labelled “unbiased”). This is expected as these stimuli are on the end-points of the stimuli.

The pattern looks rather different in Figure 5.6b. There are now non-overlapping confidence intervals in the predicted probability plot at stimulus Pair 4 (Amplified and Both) and Pair 5 (Both). Dyslexic readers categorize these pairs more often as the same than average readers do. At Pair 6

Table **5.1**
Model Evaluation With Discrimination Response (idD) As Dependent Variable. The Bayesian Deviance Information Criterion Was Used For All Consecutive Models Estimated with MCMC (150,000 iterations). D = Posterior Mean Deviance, $D(\phi)$ = Deviance of Posterior Means, $pD(D - D(\phi))$ = Effective Number of Parameters, DIC = Deviance Information Criterion. See Text For An Explanation of the Modelling Steps.

	Mulitlevel Logistic Model				
	M0	**M1**	**M2**	**M3**	**M4**
D	8041.88	7985.5	7019.02	6914.24	6789.17
$D(\phi)$	8040.8	7951.01	6965.41	6852.88	6708.45
$(D - D(\phi))$	1.08	34.49	53.6	61.36	80.72
DIC	8042.96	8019.98	7072.62	6975.6	6869.9

(Both) the dyslexic readers show a small increase in different responses compared to stimulus pair 5 and 7. The confidence intervals overlap (within the group, between the pairs as well as between the groups) and these differences are not significant. In ROC and PR space, however, Table 5.2 shows significantly different AUCs between average and dyslexic readers for Amplified and Both stimulus types. Pair 6 (Both) of the dyslexic readers has switched rank with stimulus pair 4 and in ROC space occupies the point of completely random discrimination (0.5 TPR and 0.5 FPR). In PR space there is a sudden drop in PPV after Pair 6. This pair no longer yields unbiased responses and in terms of TPR ranks between Pairs 3 and 4. The rank order switch and sudden discontinuity in the curve does not occur for average readers.

5.8.4 Conclusion & Discussion

The hypothesis under investigation was that "noise" (a noise perturbation) might yield a category switch when stimuli on a physical continuum are evaluated. This could result in the emergence of perceptual boundaries at values of the acoustic parameter where none would be perceived when the noise perturbation was absent. This effect was expected based on the 2D potential model in dyslexic readers under the assumption that their speech perception may be described by a loosely coupled system. The results provide support for the hypothesis in two ways: First, a peak in the predicted discrimination curve of dyslexic readers was observed for stimuli of type Both. The very same stimulus pair in the other manipulation conditions was the best, most unbiased discriminable stimulus pair. Second, the ROC and PR curves show discontinuities around pair 6 and the rank orders of the stimuli along the curve change when compared to the other types of acoustic manipulation. This indicates the stimulus pair changed from a reliable "same" pair into a pair that is sometimes perceived as "same" but equally often as "different".

Table **5.2**
The Area Under the Receiver Operating Characteristic Curves and Precision Recall Curves For the Four Stimulus Types. The 95% Confidence Bounds Between Parenthesis are Based on 1,500 resamples.

Manipulation	**Average readers**		**Dyslexic readers**	
	AUC (CI.95) - ROC	AUC (CI.95) - PR	AUC (CI.95) - ROC	AUC (CI.95) - PR
None	0.85 (0.83, 0.88)	0.85 (0.83, 0.88)	0.93 (0.92, 0.95)	0.93 (0.92, 0.95)
Slowed Down	0.85 (0.82, 0.88)	0.82 (0.79, 0.85)	0.93 (0.92, 0.95)	0.93 (0.91, 0.94)
Amplified*	0.85 (0.82, 0.87)	0.65 (0.61, 0.69)	0.93 (0.92, 0.95)	0.84 (0.81, 0.87)
Both*	0.85 (0.82, 0.88)	0.61 (0.57, 0.65)	0.93 (0.92, 0.95)	0.82 (0.79, 0.85)

*Non-overlapping CI.95 between average and dyslexic readers

It is unclear how these results can be interpreted in terms of an allophonic mode of speech perception. One could make the ad-hoc hypothesis, that amplification of fast changing spectral elements in the signal amplifies the native allophonic boundary, but only in interaction with the slowing down operation. Interpreting the subsequent manipulations as an increasing noise perturbation that destabilises the percepts as indicated by the ROC and PR curves and the results in Chapter 2, seems the more parsimonious interpretation.

Granted, the effects are small and in order to claim the 2D potential model underlies these results, more evidence is warranted. Experiment 2 explores the hypothesis that the internal structure of speech sound categories changes as a function of age and reading proficiency. Hypothesis 3 will be a test of the validity of the model parameters as γ_{cs}, ε and n_c will be estimated from the observations in Experiment 2.

5.9 Test 2: Internal Structure of Speech Categories is Related to Age

5.9.1 Method

Participants

The data for this study were collected over a period of two years and out of a total of 196 participants[4] 10 had to be excluded due to instrument failure. The data reported here are based on 186 Dutch children from 20 different primary schools and 2 secondary schools. The group consisted of 68 girls and 127 boys, with a mean age of 132.5 months (SE = 1.0; range = 106-181 months). The dyslexic reader group consisted of 92 children and the average or above average reader group contained 94 children. The dyslexic readers differed significantly from the average readers with regard to their reading abilities (see Table **5.3**) on two reading tasks also used in the previous experiments (the Drie-Minuten-Toets and the Klepel).

Table **5.3**
Summary statistics (Mean and SE of Mean) of the 186 Participants

	Average Reader *M (SE)*	**Dyslexic Reader** *M (SE)*
Age	132 (1.3)	133 (1.6)
DMT1	92 (2.4)	65 (2.6)
DMT2	88 (2.2)	58 (2.4)
DMT3	78 (1.9)	49 (2.0)
KLEPEL	68 (2.1)	37 (1.7)
N	94	92

Materials

As in previous studies, the stimuli were based upon natural speech recordings for the words /bAk/ [container] and /dAk/ [roof] and have been used to create a 10-step /bAk/ to /dAk/ continuum (Van Beinum et al., 2005) using the Praat program (Boersma & Weenink, 2005). In the present study however, the continuum was expanded to a 20-step continuum (using the Praat program) by manipulating not only the F2 transition onset, but the F3 transition onset as well. Table **5.4**

[4] I would like to thank the students who participated in the Research Seminar 2005 and 2006 of the School of Pedagogical and Educational Science fort heir aid in data collection.

Figure 5.5 – Predicted probabilities for perceiving a stimulus pair as "different" by average and dyslexic readers (left): None and Slowed Down. The graphs on the right show the ROC and PR curves respectively. See text for details

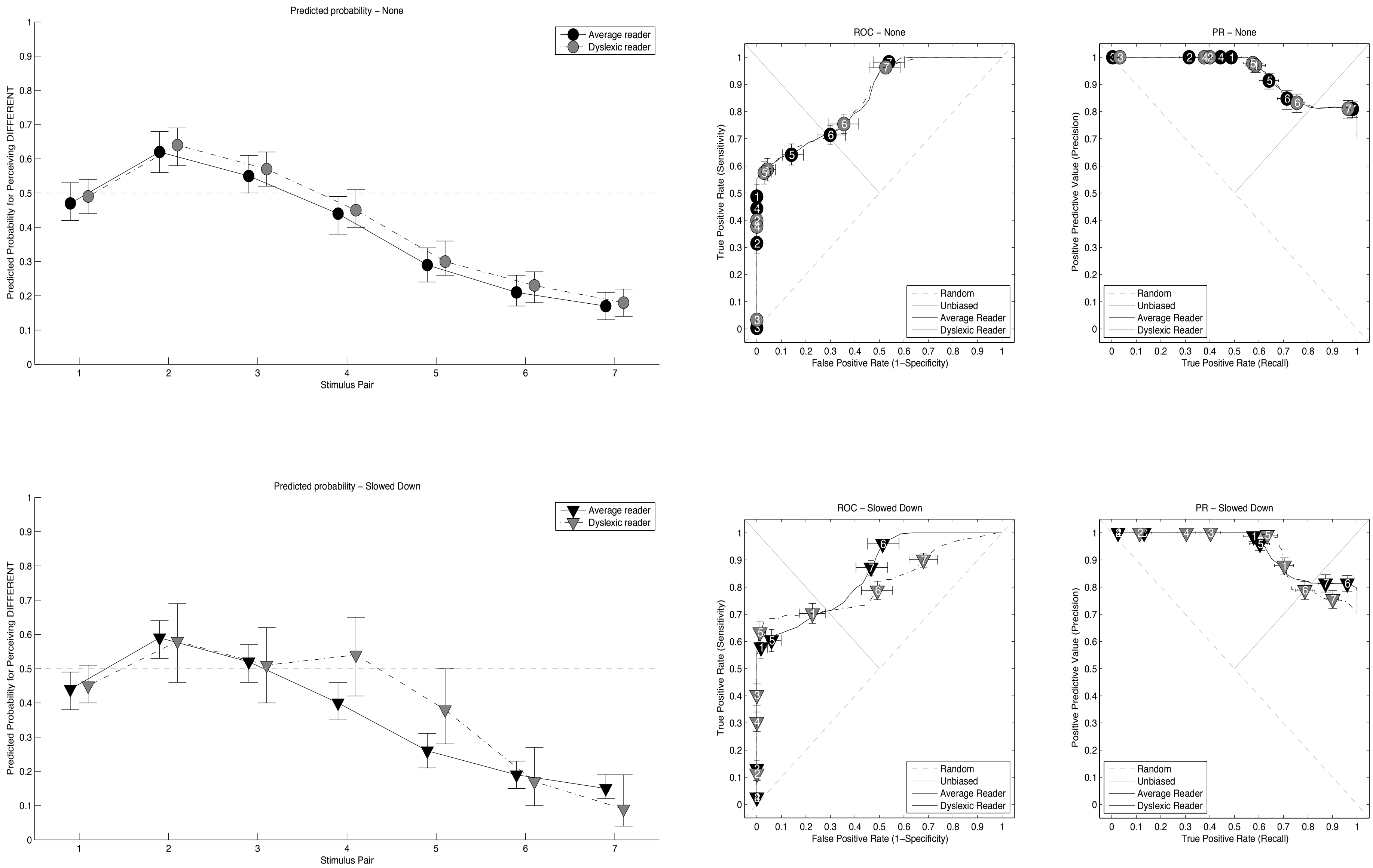


Figure 5.6 – Predicted probabilities for perceiving a stimulus pair as "different" by average and dyslexic readers (left): Amplified and Both. The graphs on the right show the ROC and PR curves respectively. See text for details

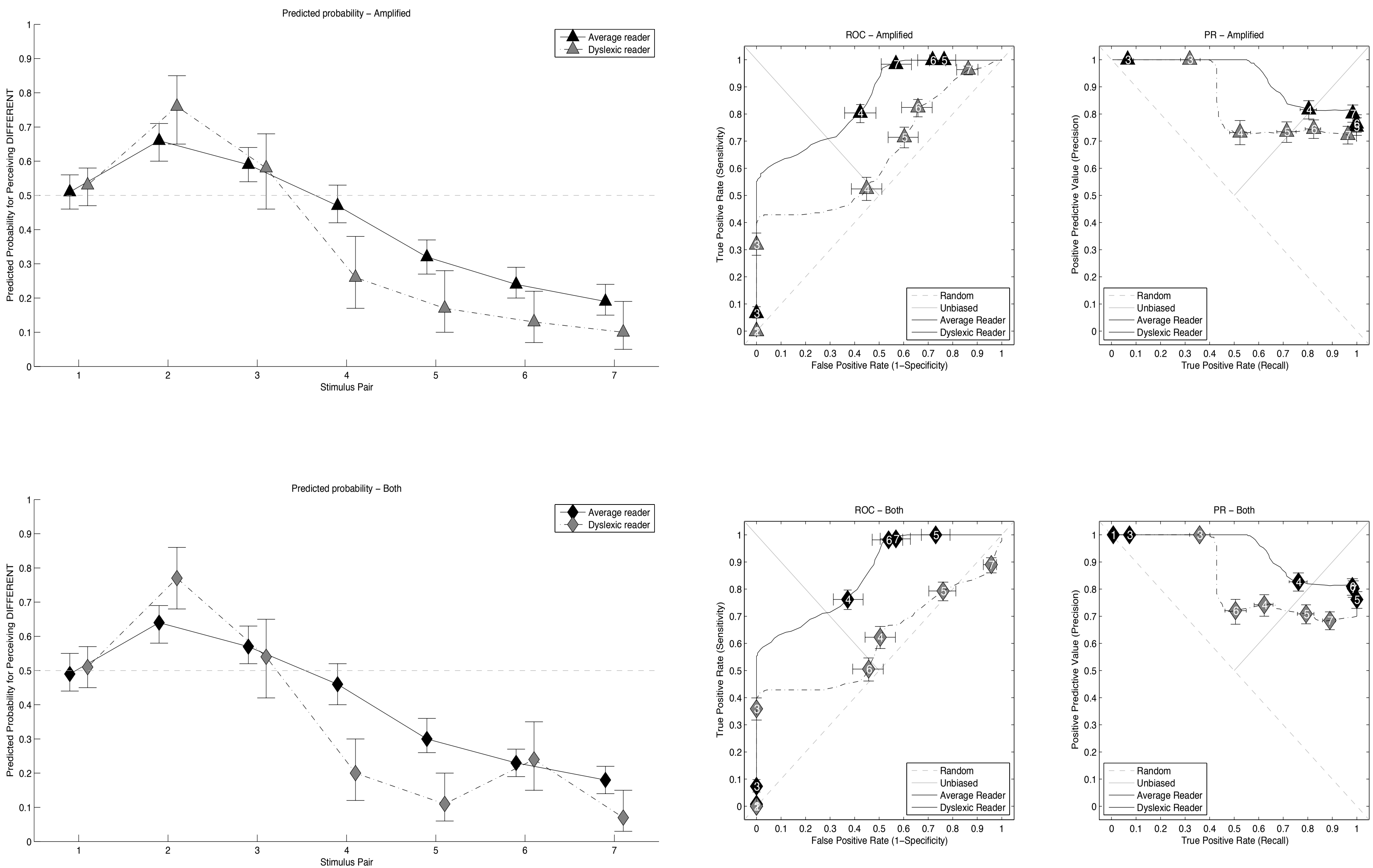


Table **5.4**
Formant Frequencies of the 20-step /bAk/-/dAk/ Continuum Used in Experiment 2.

Stimulus	F1	F2	F3	F4	F5
1	500	1080	2225	3870	4500
2	|	1117	2272	|	|
3	|	1154	2330	|	|
4	|	1191	2383	|	|
5	|	1227	2436	|	|
6	|	1264	2488	|	|
7	|	1301	2541	|	|
8	|	1338	2593	|	|
9	|	1375	2646	|	|
10	|	1412	2699	|	|
11	|	1448	2751	|	|
12	|	1485	2804	|	|
13	|	1522	2857	|	|
14	|	1559	2909	|	|
15	|	1596	2962	|	|
16	|	1633	3015	|	|
17	|	1669	3067	|	|
18	|	1706	3120	|	|
19	|	1743	3172	|	|
20	500	1780	3225	3870	4500

shows the frequencies used for the formant transition onset at each step along the continuum. The main reason for increasing the number of steps along the continuum is the increased information about internal structure of the perceptual categories that could be interpreted as “hidden” stable states, or rudimentary allophone boundaries. The decision to traverse the F2-F3 acoustic space in an oblique, rather than in an orthogonal fashion is that the allophonic perception hypothesis expects the allophone boundaries at specific frequency ranges in F2-F3 space (e.g., Noordenbos, Segers, Serniclaes, Mitterer, & Verhoeven, 2012).

5.9.2 Procedure

Speech perception experiment

The speech identification task was presented on a laptop computer in a quiet room at the children’s school and was similar to the task used in Chapter 4. The participants were presented a smiley face on the screen, which then uttered a word. After utterance of the word, two frames with a picture in each appeared on the left and right of the screen. The pictures in the frames were either a roof or a container. The pictures presented in the frames were randomly interchanged at each presentation. Prior to the experimental trials, 10 practice trials were presented using different words and pictures, which were al, clear exemplars, so the participants could receive feedback on their response. During the experimental condition, there were three different orders of presentation: (i) Sequential /bAk/ » /dAk/ » /bAk/; (ii) Sequential /dAk/ » /bAk/ » /dAk/ and (iii) Random. Each order thus consisted of 40 trials. The random order was randomised such that both halves of twenty trials contained stimuli 1-20. The orders were counterbalanced across participants, but a session always started with a sequential presentation and a sequential presentation was always followed by a random order presentation. Each order was presented twice so one participant evaluated 6 x 40 = 240 stimuli; the 20-step continuum as a whole was thus labelled 12 times by each participant.

Statistical Analysis

The hypothesis under investigation is the change of the internal structure of the perceptual categories as a function of age and reading proficiency. The participants were divided into three age groups such that the number of participants in each group was about equal. The first group ranged from 106 – 125, the second from 126 – 134 and the third from 135 – 181 months. What researchers usually present in speech perception experiments is a classification or discrimination curve like those presented in Experiment 1 of this chapter (also see Chapter 4). The procedure used to infer the existence of an allophone boundary is somewhat different and involves transforming a labelling (identification) curve into an expected discrimination curve, which is then compared to an observed discrimination curve (e.g. Bogliotti, Serniclaes, Messaoud-galusi, & Sprenger-charolles, 2008).

For the present study a different analysis strategy was developed in order to better capture the nonlinear dynamics expected by the sequential traversal of the continuum. The aim is to construct a 2D space that represents the density of transition points along the continuum, depending on the direction the continuum is traversed. A coordinate in this space with a high density could be for instance (12,9) for the sequential order /bAk/ » /dAk/ » /bAk/. This indicates many participants switch at stimulus 12 going from 1 (/bAk/) to 20 (/dAk/) and at stimulus 9 going from 20 to 1.

Parameter Space

The space is constructed in two simple steps:

1. The observed trial series of each presentation order were coded such that the starting end of the continuum was 0 and the opposite end was 1. The two random order sequences measured for each participant were sorted post hoc as if they were presented as one of the sequential orders.
2. The series were copied and (if necessary) rotated into a square matrix in such a way that one axis represented the direction of the continuum that was traversed. The series was just copied 20 times along the other axis. Each sequence order yields two matrices and thus four matrices in total were created:
   /dAk/ » /bAk/ » /dAk/
   (i) left axis, top to bottom 20->1
   (ii) bottom axis, left to right 1->20
   /bAk/ » /dAk/ » /bAk/
   (iii) top axis, left to right 1->20
   (iv) right axis, top to bottom 20->1

The matrices were summed over participants (by age group and reader group) in order to obtain the category switch density. The distributions of switch points were compared between age groups and reading groups using the Kullback-Leibler Divergence (Kullback & Leibler, 1951) or relative entropy. K-L divergence is a non-symmetric measure of the difference between two probability distributions and can be calculated according to equation 5.7

$$D_{KL}[p(x)\|q(x)] = \langle \log \frac{p(x)}{q(x)} \rangle_p = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \tag{5.7}$$

The K-L divergence can be understood as the information that is lost when one distribution is used to estimate the other. Small divergence indicates highly similar distributions, large divergence dissimilar distributions. The algorithm used to calculate the divergence does not require actual probability estimates, but computes divergence from observed values as long as all values occur in both distributions (Goni, 2007). To achieve this, the frequencies at each coordinate were transformed to the unit scale and rounded to the numbers 0.0 through 1.0 in steps of 0.1. A 95% Confidence Interval was estimated around the divergence statistic using 15,000 bootstrap replications.

5.9.3 Results

Figures **5.7** and **5.8** display the distribution densities as contour plots for each age group (3 columns) and reader group (2 rows). To understand the figure recall how they were constructed: The left y-axis and bottom x-axis represent the sequential presentation of /dAk/ to /bAk/ (y-axis top to bottom 20 » 1) and /bAk/ back to /dAk/ (x-axis left to right 1 » 20). For these trial series a /dAk/ response was coded 0 and /bAk/ was coded 1, therefore the light grey area in left lower corner represents /bAk/. The opposite coding scheme was used for /bAk/ to /dAk/ (top x-axis left to right 1 » 20) and /dAk/ back to /bAk/ (right y-axis top to bottom 20 » 1). The light grey area in the top right corner represents /dAk/ for these trial series. The iso lines connect the regions of the same density and are therefore indicative of the structure of the perceptual categories: Iso lines close together indicate a sharp transition boundary whereas interrupted or jagged iso lines represent an area which is not smoothly defined and indicate regions of greater instability. Another indicator of smoothness and stability presented only for illustration purposes are the quiver plots overlaid on the contour plots. The direction of the arrows was arbitrarily chosen to be in the direction of higher values. What is more informative to note however is whether the direction they point towards is jittered from one arrow to the next or neatly aligned. A jittered pattern is again indicative of an unstable transition region.

Evaluated qualitatively, there appears to be a tendency for the transition point densities to converge to more homogeneous areas in the older age groups. This appears to be especially the case for the average reading group. The same holds for the smoothness of the quiver directions. The K-L divergence statistics confirm this tendency as shown in Table **5.5** In order to get a sense of the significance of the divergence differences between the groups, the 95% Confidence Intervals can be compared and should be non-overlapping to give a significant difference at $p < .05$ (CIs can overlap partially as long the interquartile range is not included in the interval). The first prediction was that the density distributions of average and dyslexic readers would become more alike as a function of age (i.e. a “main” effect of age). The youngest group (1A1D = 1.81 / 2.28) has a higher divergence than the middle group (2A2D = 1.01 / 1.14), which is again higher than the older group (3A3D = 0.03 / 0.12). For the sequential presentation the relation is: A3D3 < A2D2 = A1D1 and for random presentation: A3D3 < A2D2 < A1D1. This means that the response distribution of transition points for the older age group is significantly more similar between dyslexic and average readers than between the younger groups.

The second prediction was that the older dyslexic readers would diverge less and less from the younger average readers. This effect is clearly present as A1D1 > A1D2 = A1D3 for both sequential and random presentation orders. The response distributions of the middle and older group of dyslexic readers are significantly more similar to the response distribution of the youngest average reader group. It is possible to quantify these differences by looking at the “size” of the contour area that encloses the light grey areas (representing either /bAk/ or /dAk/).

Finally, figure 5.8 displays a cross-section of the landscape along the diagonal running from the lower left corner to the upper right corner. The y-axis displays the value of the contour lines encountered at each step along the continuum. There appear to be different dips for average and dyslexic readers along the diagonal; none of these represent a significant difference between the groups.

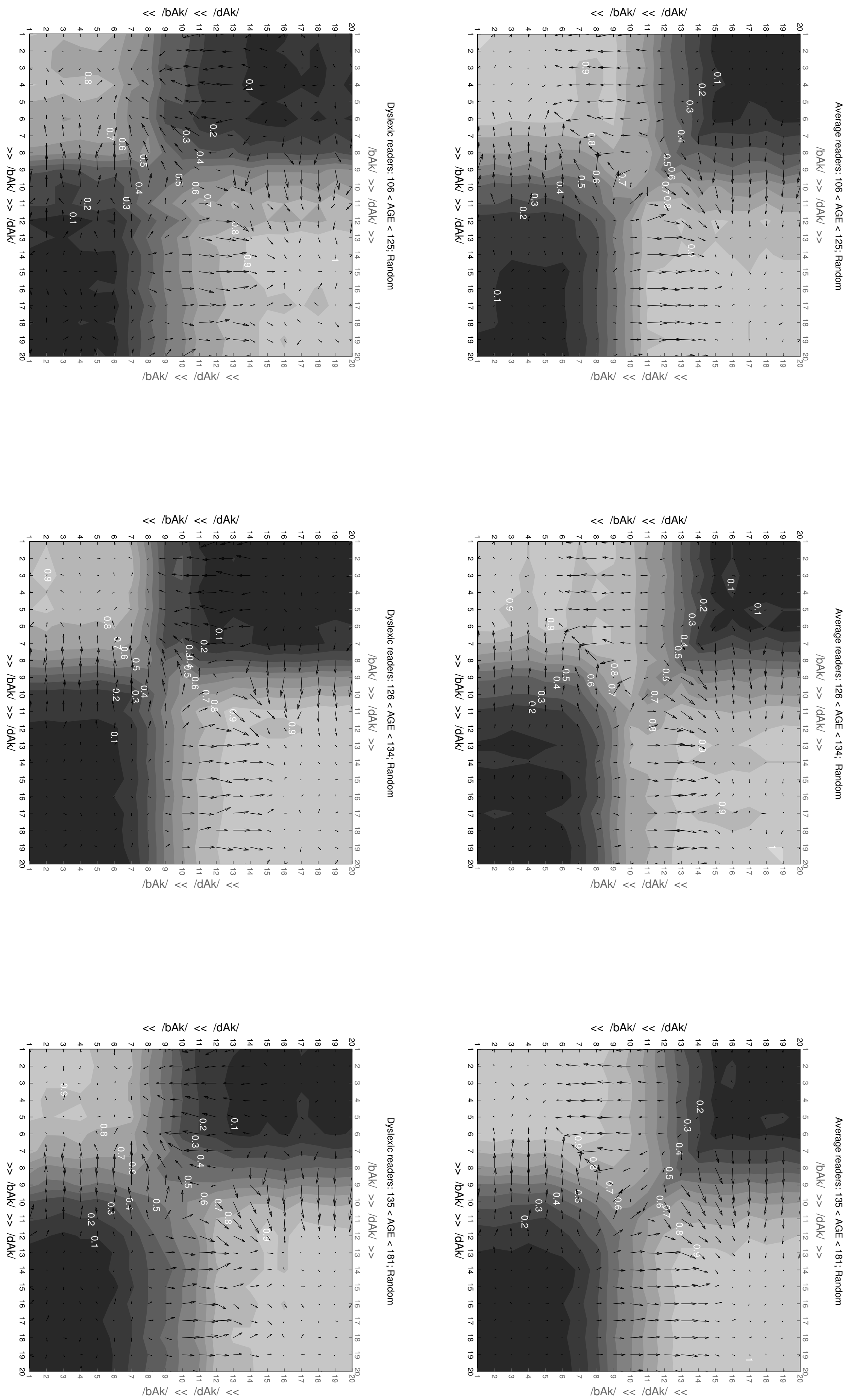


Figure 5.7 – Contour plots representing the transition point densities observed during random presentation of stimuli for each age group and reader group. The stimuli were ordered post hoc. See text for details.

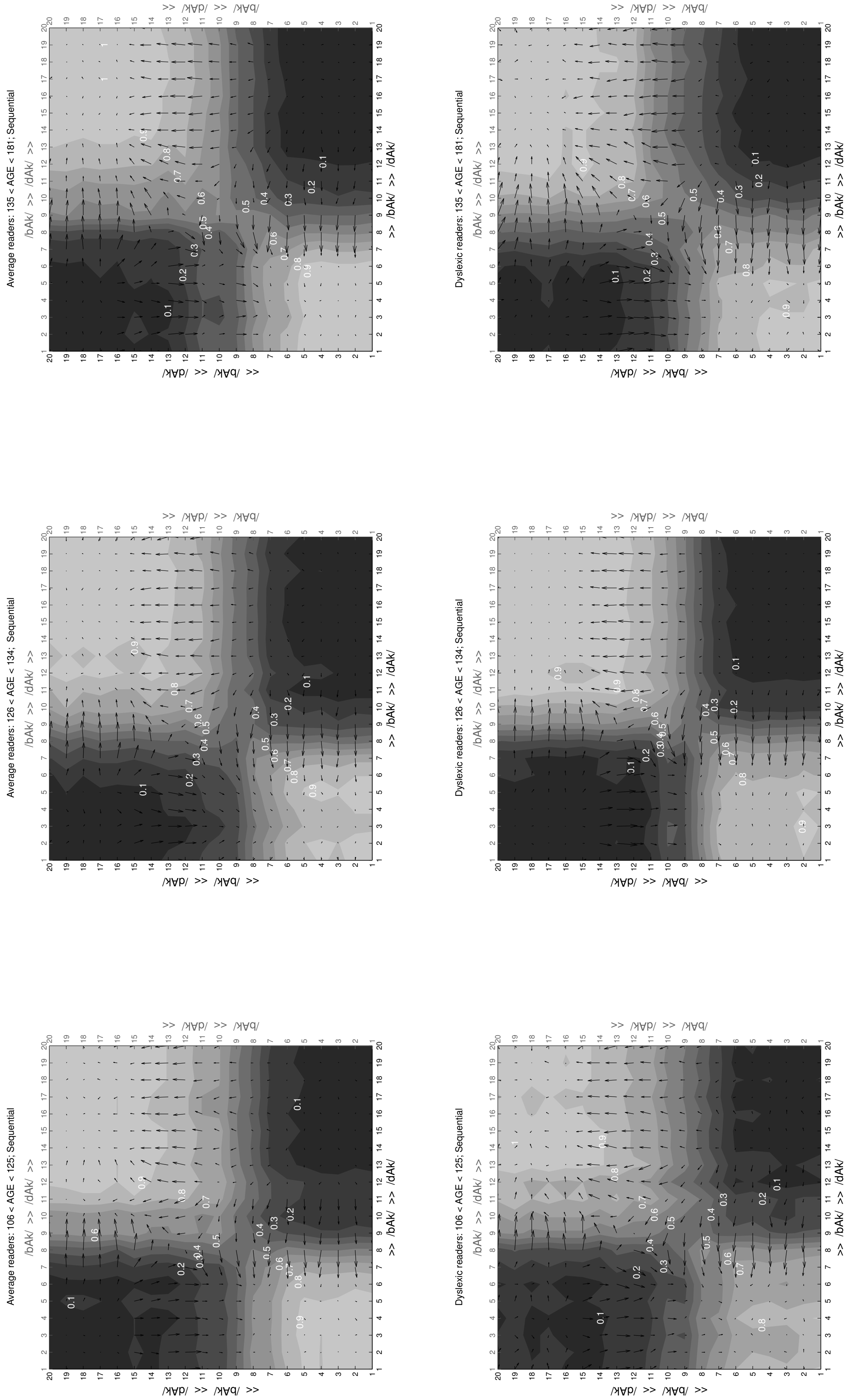


Figure 5.8 – Contour plots representing the transition point densities observed during presentation of stimuli for each age group and reader group. See text for de sequential tails.

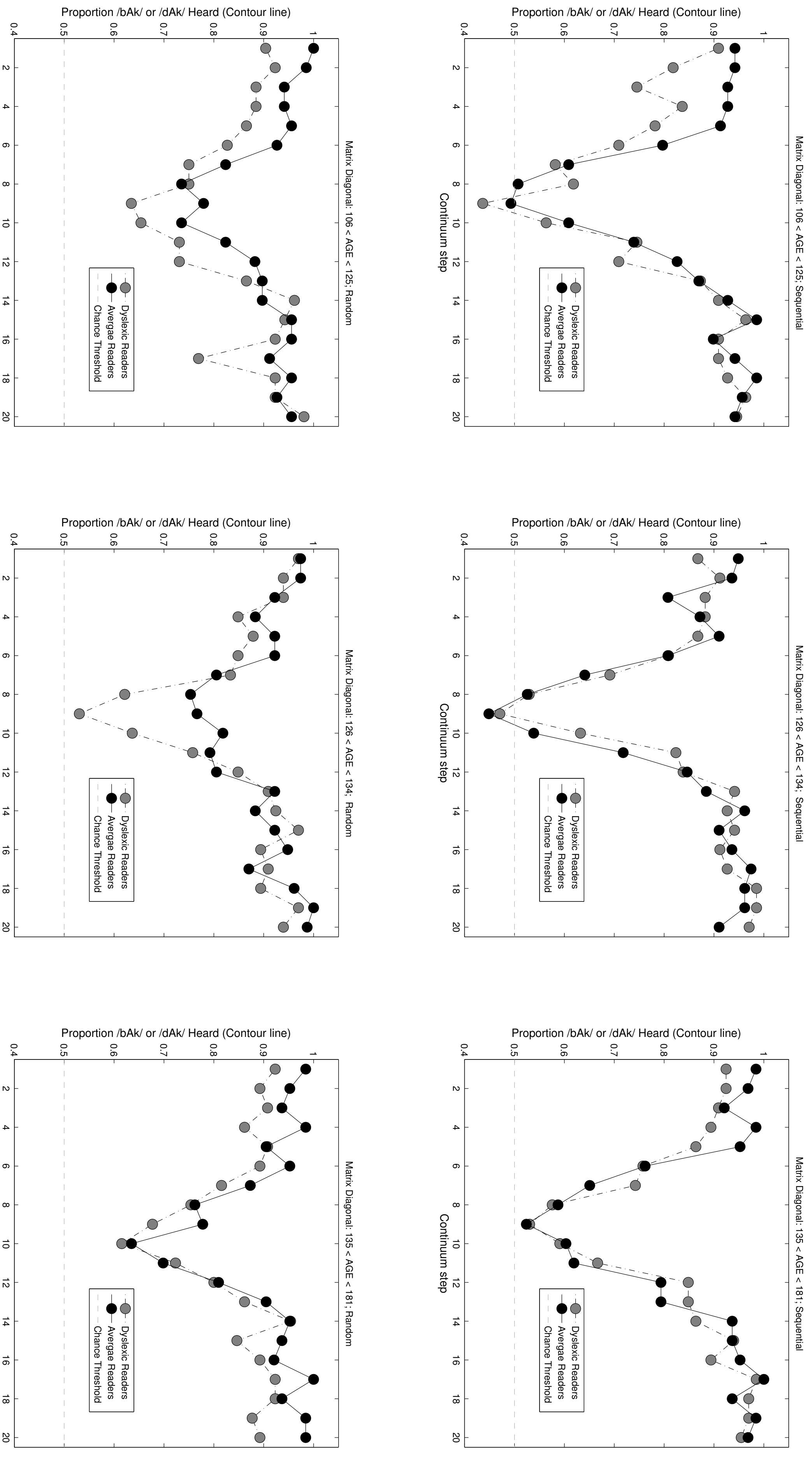


Figure 5.9 – Lines represent the values found at the contour lines (proportion of /bAk/ or /dAk/ heard) as the matrices in Figure **5.7** & **5.7** are traversed along the diagonal from lower left to upper right corner.

Table **5.5**
Formant Frequencies of the 20-step /bAk/-/dAk/ Continuum Used in Experiment 2.

	K-L Divergence (CI.95)		
	Sequential presentation		
	D1	D2	D3
A1	1.81 (1.40, 2.55)	0.03 (0.01, 0.05)	0.01 (0.00, 0.03)
A2	2.02 (1.79, 2.62)	1.01 (0.73, 1.60)	0.97 (0.69, 1.48)
A3	1.15 (0.85, 1.77)	0.08 (0.05, 0.11)	0.03 (0.01, 0.06)
	Random presentation		
	D1	D2	D3
A1	2.28 (1.79, 2.99)	0.02 (0.01, 0.03)	0.08 (0.04, 0.13)
A2	2.74 (2.49, 3.43)	1.14 (0.82, 1.68)	1.18 (0.86, 1.72)
A3	1.72 (1.32, 2.37)	0.04 (0.01, 0.08)	0.12 (0.06, 0.20)

5.9.4 Conclusion & Discussion

The aim of Experiment 2 was to investigate the age dependence of the internal structure of perceptual categories as predicted by the coupling hypotheses under investigation. It was shown that such an internal structure indeed exists and somewhat resembles a potential landscape when it is constructed as shown in Figure **5.7**and **5.8**. More importantly the structure reveals differences between average and dyslexic readers in accordance with the predictions.

The age effect was previously established in kindergarteners at risk for developmental dyslexia the context of the allophonic mode of perception (see Noordenbos et al., 2012). An important difference with the present study is that the allophonic mode of speech perception appeared to have been replaced by a phonemic mode of speech perception (i.e. categorical perception) in first grade, due to reading instruction. The age groups presented here represent children who are much older and still, the differences in their labelling of the 20-step continuum could be quantified. Proponents of the allophonic mode of speech perception may object that no evidence of actual 'additional' perceptual boundaries was presented, and they would be right. One might embark on an endeavour to quantify the width of the contour bands or contour areas in order to find such boundaries, but that it not the intention of this chapter. As said earlier, the loose coupling hypothesis predicts instabilities of perceptual categories that may be inferred from category switches that emerge dynamically and are under experimental control (see Experiment 1; Case et al., 1995), rather than static boundaries that must be uncovered in an exploratory fashion as innate formant frequencies.

The ultimate test of the validity of the potential model is of course to estimate its parameters from the data and analyse whether these parameters are in accordance with the predictions made so far. The next section introduces a novel estimation procedure for potential models and subsequently applies this procedure to the data presented in this section.

5.10 Test 3: Parameter Estimates Must Confirm Predictions

5.10.1 Method

Participants, Materials, Stimuli

The data obtained in Experiment 2 were used.

Estimating model parameters from data: A time-dependent Hidden Markov Model

Estimating model parameters like coupling strength of two potential models from observed sequences of /bAk/ and /dAk/ responses is not straightforward, nor readily available as a fully developed modelling technique. A noteworthy exception is the stochastic version of the Cusp Catastrophe that can be fitted to observed data by a maximum likelihood procedure (Grasman, Van der Maas, & Wagenmakers, 2007). The framework of Catastrophe theory is tightly related to Potential theory, but the 2D potential model presented here is in a way a "double" cusp, or a multivariate catastrophe of the umbilical kind. The current form does not exist as a known Catastrophe. Moreover, the parameters ε and n_c that need to be estimated are not easily translated to "regression"-like weights like k, γ_{cs} and the "fixed" parameters (1/2 and 1/10, see eq. 5.6).

In this section I will therefore propose a novel method that combines two well-known frameworks: Hidden Markov Models (HMM) and so called brute force Maximum Likelihood Estimation (bf-MLE). The HMM approach is used to model observed sequences of binary or multinomial random variables in which the observations are considered emissions generated by unobservable (hidden) states (cf., Durbin, Eddy, Krogh, & Mitchison, 1998). Suppose we use a biased coin to decide to roll 1 die on heads (80% of tosses) or roll 2 dice on tails (20% tosses). Then the hidden states of this system are Heads and Tails and its observable emissions are 1 through 12. The transition matrix T contains the probabilities of changing from heads to tails at each toss and the emission matrix E contains the probabilities of an observed emission being generated by a hidden state. In this example T and E are:

$$T = \begin{bmatrix} P_{H\to H} & P_{H\to T} \\ P_{T\to H} & P_{T\to T} \end{bmatrix} = \begin{bmatrix} 0.80 & 0.20 \\ 0.80 & 0.20 \end{bmatrix}$$

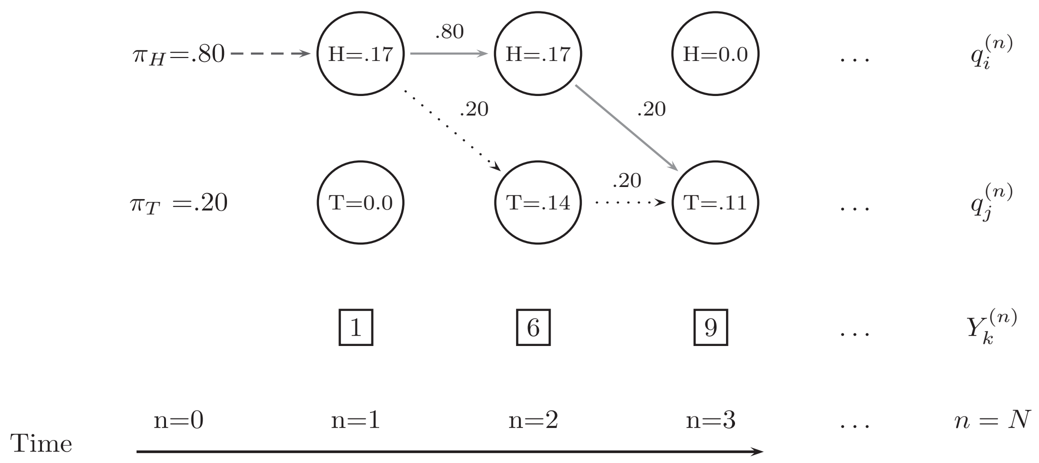
$$E = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/36 & 2/36 & 3/36 & 4/36 & 5/36 & 6/36 & 5/36 & 4/36 & 3/36 & 2/36 & 1/36 \end{bmatrix}$$

Suppose an observed sequence Y consists of three emissions $\{1, 6, 9\}$, then the posterior state probabilities of each emission in the sequence are:

$$\lambda = \begin{bmatrix} 1 & 0.83 & 0 \\ 0 & 0.17 & 1 \end{bmatrix}$$

Based on these probabilities the most probable path through this HMM resulting in the observed emissions is the state sequence $\{H, H, T\}$. There is also a less probable path through states $\{H, T, T\}$. Figure **??** shows the paths through the HMM as a Trellis diagram and reports how to calculate the likelihood of the sequence given a specific path (1), or all possible paths (2) through the HMM with parameter set $Lambda = \{\pi, T, E\}$.

An HMM is memory-less in the sense that previous states do not influence the probabilities of future states and this is of course quite contrary to the purpose of the potential model presented in this chapter. The framework of transitions between hidden states that can generate emissions however appears rather appropriate: There are hidden stable states, wells in the potential landscape that can generate one of two emissions: /bAk/ or /dAk/. The difference with the HMM is that the potential landscape changes as the k parameter changes with each stimulus presentation. This implies that a transition and emission matrix should be constructed for each stimulus presentation that somehow reflect the probability of a transition from one state to another based on a previous trial and the probability of a response being emitted by a state given the current trial. Defined as such this model may be described as an HMM with time-dependent emission and probability matrices. Box **5.2** gives a mathematical description of the model setup as a simulation of the present experiment in which 40 trials of a 20-step /bAk/-/dAk/ continuum are presented in sequential or random order to average and dyslexic readers.



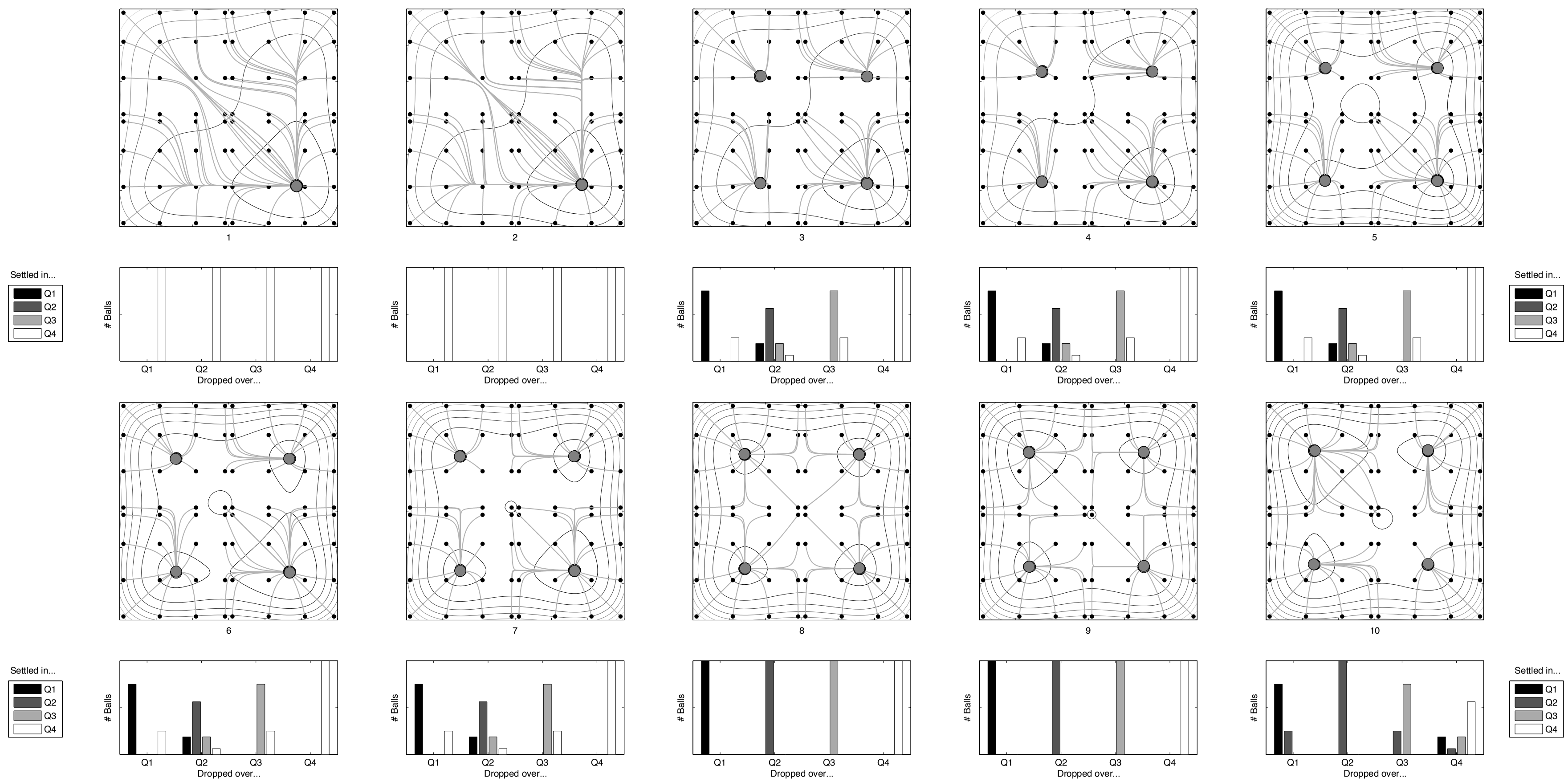
1. Joint likelihood of $Y_k^{(n)}$ and a path $Q = \{q_{i \to j}^{(n)}\}$ (e.g. dotted lines) through HMM (Λ):
   $\mathfrak{L} = P(Y,Q|\Lambda) = P(Q|\Lambda) * P(Y|Q,\Lambda)$ (Bayes)
   $\mathfrak{L} = (.80) * (.20 * .20) * (.17 * .14 * .11) = .00008$
2. Likelihood of $Y_k^{(n)}$ given this HMM with parameters Λ:
   $\mathfrak{L} = \sum_{all\ Q} P(Y,Q|\Lambda)$
   $\mathfrak{L} = (.80) * (.20 * .20) * (.17 * .14 * .11) + (.80) * (.80 * .20) * (.17 * .17 * .11) = .0005$

Figure 5.10 – A Trellis diagram of the HMM in which the outcomes of a biased coin toss represent the hidden states of the model (q_i). The states are associated with observable emissions Y_k in this example the numbers 1 to 12. The diagram shows the prior probabilities p_i (at $n = 0$), the probability that a state emitted the observation Y_k at time=n and the state transition probabilities (above the arrows).

Assume we divide the potential landscape into four quadrants that represent four possible hidden states q_i. More precisely, a hidden state q_i is defined as the quadrant of the potential landscape where an imaginary ball settle, should it be dropped from an arbitrary location above the landscape. The transition probabilities from state q_1 to state q_2 can then be defined as the number of balls ending up in q_2 that were initially released over q_1. If q_2 happens to be a deep potential well, for instance in the case of $k = -1$ or $k = 1$, all the balls dropped over q_1 would end up in q_2 and the probability to switch from q_1 to q_2 would be 1. For other parameter settings things could be very different, for instance when $k = k_c$ the balls dropped over q_1 are likely to remain in q_1 and a transition to another state has a low probability. The transition matrix is thus a 4x4 matrix, its rows represent the quadrant the balls were dropped from and its columns represent the quadrants the balls end up in. In the first situation sketched above cell 1,2 of the matrix would contain a 1, denoting that all balls dropped above q_1 end up in q_2. In the second situation cell 1,2 would probably contain a 0, indicating none of the balls dropped in q_1 ended up in q_2. In the estimation procedure a total of 16 balls will be dropped above each quadrant, 64 in total.

Figure **5.11** is an example of this procedure and contains the two situations just discussed. The figure shows 10 trials that are simulated for parameter settings γ_{cs} = 0, ε = 0.5 and n_c = 10, the simulation starts at k = -1. Shown are the potential landscapes with drop points and the path an imaginary ball would follow through the landscape were it dropped at that point. The bar graph below the landscape displays where balls dropped over each quadrant eventually settle. For the first stimulus the result is very clear and resemble the first situation described above: All balls end up in q_4 which we associate with either /bAk/ or /dAk/ as the end-points of the continuum. Stimuli 8

Figure 5.11 – An example of 10 simulated trials for model parameters values $\varepsilon = 0.5$, $\gamma_{cs} = 0$ and $n_c = 10$. The black dots above the potential landscape indicate drop points from which imaginary balls are dropped. The grey trajectories display the path a ball traverses through the landscape; the large grey circles indicate the endpoints, i.e where a ball settles after it was dropped. The bar graphs indicate for the drop points over each quadrant in which quadrant balls eventually settle. See text for details.



Box 5.2: Definition of the Hidden Markov Model with time-dependent transition and emission probabilities used to simulate the experiments. See text for details.

$$\textit{Given an observed response sequence } Y_k^{(t)}, \textit{ and simulated } \Lambda_{\langle\varepsilon,n_c,\gamma_{CS}\rangle}:$$

$$\mathfrak{L}\left(\Lambda_{\langle\varepsilon,n_c,\gamma_{CS}\rangle}\|Y_k^{(t)}\right) = \prod_{t=1}^{|Y^{(t)}|} f\left(Y_k^{(t)}\|\Lambda_{\langle\varepsilon,n_c,\gamma_{CS}\rangle}\right)$$

$$let \begin{cases} N &= 40 \\ \varepsilon &= -.5,\ldots,.5 \\ n_c &= 10,\ldots,30 \\ \gamma_{CS} &= 0,\ldots,1 \\ \Lambda_{\langle\varepsilon,n_c,\gamma_{CS}\rangle} &= \{\Phi,\Theta,\Pi\} \\ \lambda_\pi^{(t=1)} &= P\left(\theta^{(t=1)} = \theta_\pi^{(t=1)}|Y_k^{(t=1)}\right) = \pi_i \\ \lambda_i^{(t)} &= P\left(\theta^{(t)} = \theta_i^{(t)}|Y_k^{(t)}\right) \\ \lambda_j^{(t+1)} &= P\left(\theta^{(t+1)} = \theta_j^{(t+1)}|Y_k^{(t+1)}\right) \end{cases}$$

$$\begin{aligned} &Assume: \\ &\forall\, Y_k^{(t)} \in \{/bAk/,/dAk/\} : V(x,y)_{\langle\varepsilon,n_c,\gamma_{CS}\rangle} \vdash \lambda_{\pi ij}^{(t)} \therefore \\ &\theta^{(t)} = \operatorname*{arg\,max}_{\pi ij} \lambda_{\pi ij}^{(t)} \end{aligned}$$

$$\textit{Then the Maximum Likelihood Estimate of the tuple } \langle\varepsilon,n_c,\gamma_{CS}\rangle \textit{ given } Y_k^{(t)} \textit{ is defined as}:$$

$$\begin{aligned} \mathfrak{L}\left(\Lambda_{\langle\varepsilon,n_c,\gamma_{CS}\rangle}\|Y_k^{(1)},...,Y_k^{(N)}\right) &= \prod_{m=1}^{N-1}\prod_{n=2}^{N}\left(\phi_{\lambda_j^{(m)}\to\lambda_j^{(n)}}\,\theta_{\lambda_j}^{(n)}\,\pi_i\right) \\ \ln\,\mathfrak{L}\left(\Lambda_{\langle\varepsilon,n_c,\gamma_{CS}\rangle}\|Y_k^{(1)},...,Y_k^{(N)}\right) &= \sum_{m=1}^{N-1}\sum_{n=2}^{N}\ln\left(\phi_{\lambda_j^{(m)}\to\lambda_j^{(n)}}\,\theta_{\lambda_j}^{(n)}\,\pi_i\right) \therefore \\ \langle\varepsilon,n_c,\gamma_{CS}\rangle_{MLE} &= \operatorname*{arg\,max}_{\varepsilon,n_c,\gamma_{CS}} \ln\,\mathfrak{L}\left(\Lambda_{\langle\varepsilon,n_c,\gamma_{CS}\rangle}\|Y_k^{(1)},...,Y_k^{(N)}\right) \\ \langle\varepsilon,n_c,\gamma_{CS}\rangle_{MLE} &= \operatorname*{arg\,max}_{\varepsilon,n_c,\gamma_{CS}} \sum_{n=1}^{N} \ln\,P\left(Y_k^{(1)},...,Y_k^{(N)}|\Lambda_{\langle\varepsilon,n_c,\gamma_{CS}\rangle}\right) \end{aligned}$$

and 9 show a different pattern, akin to the second example described above: now all the balls settle in the same quadrant they were dropped over. These parameter settings should yield Enhanced Contrast behaviour (because ε = 0.5), which is indeed likely as the multi-stable landscape presents itself early in the trial series.

Apart from transition probabilities there are also emission probabilities that link the transition matrix to actual observations. The emission matrix in the present case is 4 rows by 2 columns; the latter represent the observable states (/bAk/ and /dAk/). Its rows represent the probability that the /bAk/ or /dAk/ can be observed, given that the model is in the state represented by the row (q_1 to q_4). Due to the way the potential model is constructed, there are two hidden states that will represent either the observable sates most often. These quadrants are q_2 and q_4, the only stable minima that exist at the extremes of the continuum. The other two hidden states are the spurious states, by-products of the coupling of the two potential landscapes. It is however possible, and this was the reason for choosing this estimation scheme, that a switch from /bAk/ to /dAk/ was caused by hidden state q_3 or q_1. In the case of stimulus 1 in Figure **5.11**, the entire second column will contain a 1 because no matter in which state the model is, we will always observe emission 2. For stimulus 8 in the same figure the emission matrix looks a bit different. Given that the system is in q_2 or q_4 the likelihood of observing the associated observation is 1. However, by definition q_1 and q_3 cannot emit an observable state by themselves, so the other cells in the matrix are 0. Another way to say this is that at each trial the model's initial state probability is always the observed label on the previous trial and q_1 and q_3 are not associated with any category labels. The only occasions in which q_1 and q_3 appear with probabilities in the emission matrix are when a ball was dropped over q_2 or q_4 and settled in q_1 or q_3. This signals a transition of states and perhaps of perceived category, as is for instance the case with stimulus 10 in Figure **5.11**. Several balls dropped over q_4 end up in q_1 and q_3, this contributes to the likelihood of observing emission state associated with q_2.

The clue to adding a "memory" to this HMM model is pretty straightforward. As the initial state probabilities we can use the probabilities associated with previously observed emission. Obviously this can only be done for trials > 1 and the starting state probabilities just represent the first observation at probability 1. There is one major difference to the context of regular HMM model fitting that makes this possible; the potential model can be simulated for a large range of parameter settings so that the matrices for each trial sequence, for each parameter setting are known in advance. The algorithm enumerates through all the simulated parameter values and calculates the likelihood of an observed trial series using the transition and emission matrices. After all the parameter settings have been evaluated, the parameter tuple that yielded the highest likelihood for the trial series is selected. Box **5.3** explains how the maximum likelihood of an observed sequence was obtained (Myung, 2003 is a tutorial).

5.10.2 Procedure

Parameter Estimation and Hypotheses

The potential model was simulated using Matlab R2012a (The MathWorks, 2012). First let's consider which parameters to include in the simulation. The value of parameter ε governs the slope of the potential landscape through its effect on k and therefore is decisive in observing hysteresis or enhanced contrast behaviour (cf. Tuller et al., 1994). The parameter n_c affects at which stimulus the effect of ε on the slope of the potential becomes noticeable and is therefore indicative of the size of the difference between category switches up and down the continuum. If we were to estimate "just" the parameter ε and fix n_c at 20 (i.e. after which the continuum is traversed in the opposite direction) valuable information might be lost. Therefore ε was estimated as a restricted parameter taking on three values in order to grossly capture the nonlinear behaviour: at -.5 we observe hysteresis, at 0 we observe something akin to a critical boundary at .5 we observe enhanced contrast. The parameter n_c was fitted for values 10 to 30 in steps of 1 (21 values). If $\varepsilon = 0$, the value of n_c cannot be estimated because after n_c there is no change in slope by the extra term $\varepsilon * (\lambda - \lambda_f)$ added to k (see eq. refeq53). So for $\varepsilon = 0$, n_c was set fixed to 20. The coupling strength γ_{cs} was simulated for parameter values 0 (no coupling) to 1 (full coupling) in steps of .05 (21 parameter values). For these parameter ranges the experiment needs to be simulated 2x21x21 + 1x21x1 = 903 times. Each simulation starts at $k = -1$ and concerns 40 stimuli of a 20 step continuum. The 20 step values for λ representing the location in F2-F3 space are linearly equivalent to the values reported in Table Step over a range of 0 to 2 (with λ_f fixed at 2). These values for λ give a value for the slope that varies from $k_\lambda = 0, n = 1 = -1$ to $k_\lambda = 2$,$n = 20 = 0$ to $k_\lambda = 0$, $n = 40 = -1$. For other parameter settings the value of k depends on ε and n_c.

For each of the 903 parameter combinations, the transition and emission matrices of 40 trials needed to be calculated. A total of 36,120 potential landscapes were generated). The parameters obtained by maximum likelihood for each observed trial series were subsequently analysed in a linear mixed effects model using γ_{cs} and n_c in separate analyses as the dependent variable. In both models ε (as a three level factor), age (as a continuous covariate), presentation order (as a three

Box 5.3: Definition of the Maximum Likelihood Estimator used to estimate parameters ε and γ_{cs} of the 2D potential model for each observed experimental trial sequence. See text for details.

$$Number\ of\ simulated\ stimuli\ -\ N = \{40\}$$
$$Trial\ indices\ -\ m = 1, \ldots, (N-1)\ and\ n = 2, \ldots, N$$

$$Hidden\ states\ -\ Q = \{q_i\},\ i = 1, \ldots, 4\ :$$
$$q_i \in \{\text{I(+,+),II(-,+),III(-,-),IV(+,-)}\}\ of\ V(x,y)_{\langle \varepsilon, n_c, \gamma_{CS} \rangle}$$
$$Transition\ probabilities\ -\ \Phi = \left\{ \phi_{i \to j} = P(q_j^{(n)} | q_i^{(m)}) \right\}$$

$$\Phi_{q_i} \underrightarrow{(m,n)}_{q_j} = \begin{bmatrix} P_{q_1,q_1}^{(m,n)} & \cdots & P_{q_1,q_j}^{(m,n)} \\ \vdots & \ddots & \vdots \\ P_{q_i,q_1}^{(m,n)} & \cdots & P_{q_i,q_j}^{(m,n)} \end{bmatrix} : \sum_{q_j=1}^{|Q|} P_{q_i,q_j}^{(m,n)} = 1$$

$$Observable\ emissions\ -\ O = \{o_k\},\ k = 1, \ldots, 2\ :\ o_k \in \{/bAk/, /dAk/\}$$
$$Emission\ probabilities\ -\ \Theta = \left\{ \theta_{jk} = \theta_j(o_k) = P(o_k^{(n)} | q_j^{(n)}) \right\}$$

$$\Theta_{o_k,q_j}^{(n)} = \begin{bmatrix} P_{o_1,q_1}^{(n)} & \cdots & P_{o_k,q_1}^{(n)} \\ \vdots & \ddots & \vdots \\ P_{o_1,q_j}^{(n)} & \cdots & P_{o_k,q_j}^{(n)} \end{bmatrix} : \sum_{o_k=1}^{|O|} P_{o_k,q_j}^{(n)} = 1$$

$$Initial\ state\ probabilities\ -\ \Pi = \{\pi_i = P(q_i^{(m=1)})\}$$

$$Time-dependent\ HMM\ -\ \Lambda = \{\Phi, \Theta, \Pi\}$$

level factor) and reading group (as a two level factor) were used as predictors. The random effects structure contained two unrelated sources of random variation around the intercept: participant ID and response label (0-1) nested within stimulus number (1-20). To fit these models the R statistical software (R Core Team, 2012) was used with the nlme package for (non)linear mixed effect models (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2012) and the multcomp package (Hothorn, Bretz, & Westfall, 2008) to conduct the multiple comparisons after lme modelling. The general strategy for modelling experiments with crossed random effects was followed (see Baayen, Davidson, & Bates, 2008 for details).

The most important hypothesis to be tested is the coupling hypothesis: Examine whether the fitted values for γ_{cs} are related to age and reading group. Based on the model, but also on the results from Experiments 1 and 2, the average readers were expected to show a stronger positive relationship between γ_{cs} and age than the dyslexic readers. Also, coupling strength was considered as a structural variable that should be rather consistent within a participant and within levels of age. More specific, the general effect of age on coupling strength should not be very different when conditioned on covariates like the order of presentation or the type of nonlinear behaviour.

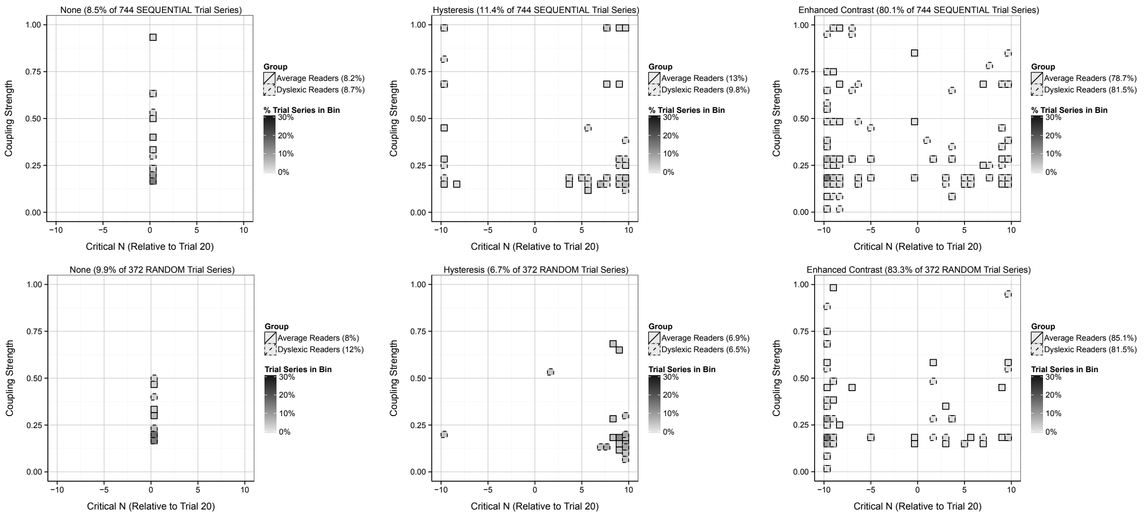


Figure 5.12 – The results of parameter estimation. Parameters γ_{cs} and n_c are shown for levels of ε in separate figures. The top row pertains to sequentially presented trial series, the bottom row to series presented in random order. The bins are colour coded to represent the frequency of occurrence of a certain γ_{cs}, n_c tuple. See text for details.

If the parameters for coupling strength would greatly differ in range when evaluated within these groups, this would pose a serious problem for the model assumptions so far.

The critical number of stimuli n_c beyond which the effects of ε start to dominate the behaviour of the model by adjusting the slope of the potential landscape, is expected to vary within a participant, probably even taking on rather different values for each sequence of trials observed within one participant. Especially when n_c would be conditioned on the covariates for which γ_{cs} is supposed to stay rather constant. More specifically: A relation between age and n_c is expected for sequentially presented stimuli, but with a different sign of slope for Hysteresis and Enhanced Contrast. Trial series classified as Hysteresis (ε = .05) should show a high n_c at a young age signifying a large gap along the continuum for the delayed switch. The size of this gap should decrease with age under influence of the increasing coupling strength. The opposite pattern should be observed for sequential trial series classified as Enhanced Contrast ($\varepsilon = -.05$). Now n_c should start at a low value and should increase with age. The general expected effect of reading group is that irrespective of the sign of the slope, the average readers slopes are larger. In the case of no bias ($\varepsilon = 0$) there should be no relation between n_c and age as this value was fixed at 20. Due to the random effects in the lme model, we could observe some fluctuations around this value.

For trial series presented in random order there is, contrary to the case of γ_{cs} estimation, no reason to expect a relation between n_c and age. The random order series were all put back in sequential order before they were submitted to the fitting procedure. There are a large number of stimuli along the continuum (20) and the randomisation was done within each half of the 40 trial sequences, so nonlinear categorisation behaviour is indeed expected. In principle though, any nonlinear categorisation behaviour should be regarded as partially accidental, partially due to random participant characteristics interacting with the experimental procedure (e.g., Cox, Hasselman, & Seevinck, 2011) and partially due to γ_{cs}.

5.10.3 Results

The results of the estimation procedure are graphically displayed in Figure **5.12**. Each bin in the graphs represents a (γ_{cs}, n_c) parameter tuple that was observed at least once, the colour codes the frequency with which the tuple occurs. Overall, a wide range of parameter values was estimated.

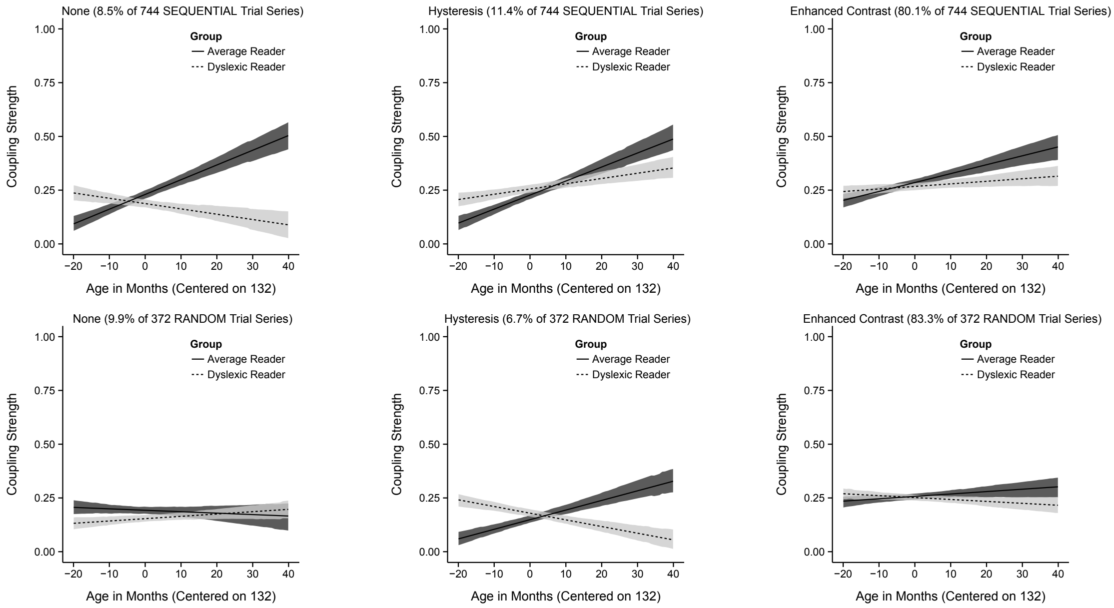


Figure 5.13 – The most relevant interaction effects predicted for Coupling Strength by the model summarised in Table The two sequential orders used in the experiment were merged to present the effects for all sequential order trial series in the data set (top row).

An unexpected result is the large difference in amount of trial series estimated for different levels of ε: The majority of trial series are estimated as ε = 0.5. What was expected is that the trial series estimated to show Hysteresis (ε = - 0.5) appear to coincide with positive n_c, or 'late' jumps. Enhanced Contrast (ε = 0.5) more frequently co-occurs with negative values for n_c indicating 'early' jumps. Remember though that n_c does not designate the actual category switch, it indicates when the influence of ε will be noticeable.

The results of the interactions predicted by the linear mixed effect model with coupling strength as a dependent variable and age, reading group, presentation order and ε as covariates are shown in Table 5.7. The 95% Highest Posterior Density Intervals (also known as the Bayesian Credible Interval), together with the covariate estimates (MCMC mean) and posterior p-values (pMCMC) were obtained using 15,000 MCMC replications. In Figure **5.13** the model predictions of the most important interactions are displayed. For presentation convenience, the two sequential orders of presentation were merged together and are displayed as a separate SEQUENTIAL subset of Trial Series (top row of Figure **5.13**). The coloured bands around the prediction lines represent the 95% HPD interval estimated from the posterior MCMC distribution. Analogously to γ_{cs} Table 5.8 and Figure **5.14** show the results of the lme model fit with n_c as the dependent variable.

It is of course possible that the fit results were obtained purely based on the dynamics between the two most stable states. Figure **??** shows the posterior state probabilities (log transformed and summed) for the three age groups (A/D1, A/D2, A/D3) as used in Experiment 2. The figures show participant group (rows) and presentation order (columns) for each age group. The posterior probabilities were obtained for each trial using the results from the fitting procedure (i.e. the most likely parameters). Each graph shows the 40 trials of each experiment on the x-axis and the rows represent states of quadrant Q2, Q1+Q3 and Q4 (see Figure **5.11**) of the potential landscape. The colouring indicates the likelihood that an observed response /bAk/ or /dAk/ was emitted from state Q1-Q4 (with Q1 and Q3 representing the spurious states). What can be seen is that as age increases, the row of the spurious states becomes lighter especially for the average readers. This indicates the observed trial series are less likely to have been generated by a hidden state for older participants, whereas for the dyslexic readers this likelihood is higher. This is in accordance with the reported effects of increase in coupling strength with age, which should lead to less relative stability of the

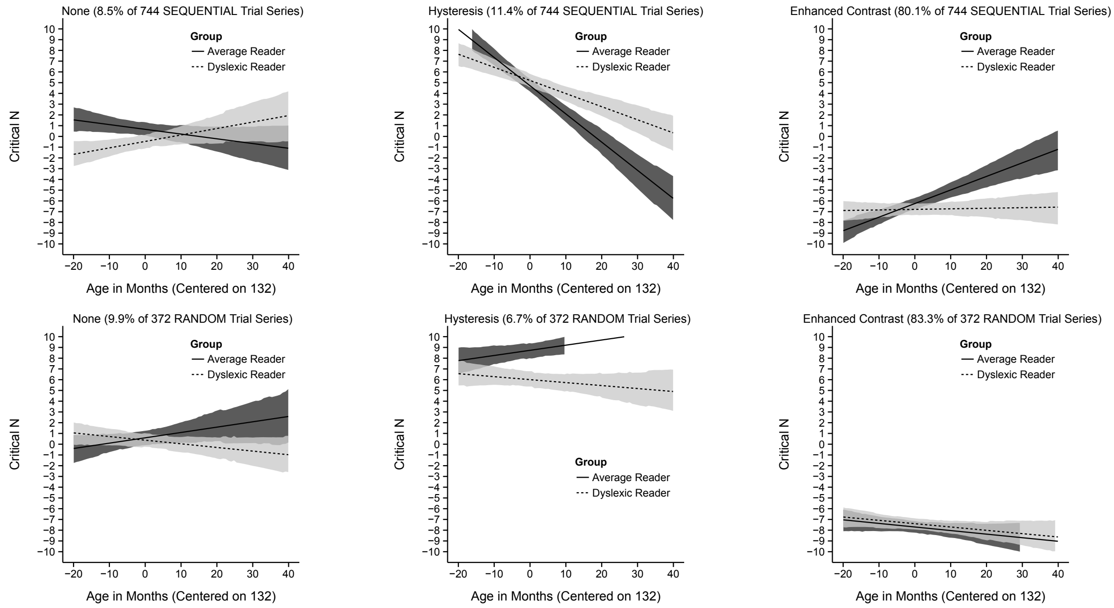


Figure 5.14 – The most relevant interaction effects predicted for Coupling Strength by the model summarised in Table 5.7. The two sequential orders used in the experiment were merged to present the effects for all sequential order trial series in the data set (top row).

spurious states Q1 and Q3.

This effect is quantified in Table 5.9 by correlating the likelihood for each group in the graph with the reading test outcome measures (Klepel, DMT1, DMT2, DMT3). Negative correlations indicate a relation between higher reading test performance and less likelihood that the trial series was generated based on spurious, or hidden states. Correlations are significant only for A3.

5.11 Conclusion & Discussion

The novel method for estimating the parameters of a 2D potential model introduced in this section was successful on several accounts: 1) The time dependent HMM framework yielded unambiguous maximum likelihood estimates for the observed trial series. 2) The obtained estimates were meaningful and in accordance with previous empirical findings and predictions based on the coupling hypothesis.

The most important finding was the confirmation that age is positively associated with coupling strength, and that this association is stronger for average readers than for dyslexic readers. A second important result is that the specific role hypothesised for the parameters in the model seems to be accurate: Coupling strength is a structural parameter, relatively constant within a participant or participant group. The other parameters do vary within participants and participant groups grossly in the direction that was expected. Figure **5.13** shows the specific predictions for the relation between age and n_c seem to be confirmed for the sequential trial series, although the Enhanced Contrast series show no association with age for dyslexic readers. These results also showed that the differences between average and dyslexic readers in the size of early or delayed category switches are expected to be lager for older children. These differences can only be revealed to exist however when presentation order is sequential, not random.

Finally, the fitted results were used to obtain state probabilities. This is an important clue to whether the hypothesised spurious states actually play a role in generating the observed sequence. When subjects were aggregated at the level of the trials in the experiment (Figure **??**) there was indeed a visible trend of decreasing likelihood of states Q1 and Q3 emitting the responses with age. This trend was corroborated when the responses of each individual participant were correlated with reading test performance. This final test provides the link to the symptoms associated with developmental dyslexia: The global trend is that a low likelihood of spurious states influencing speech perception performance is associated with age (i.e. coupling strength). In average readers this trend is statistically significant, not so for the dyslexic readers where this trend is however visible as the correlations clearly increase with age. These results appear to mimic the results from Experiment 2 where dyslexic readers lag behind with respect to the average readers increase in variables associated with coupling strength.

An unexpected result was the distribution of trial series over values for ε. The majority of trial series, irrespective of presentation order, was classified as Enhanced Contrast. The most plausible reason for this is that the values at which ε was fixed were not chosen correctly. Figure **5.13** hints this might be the case as there are a considerable number of Enhanced Contrast trial series estimated at n_c values one would expect for Hysteresis trial series. The reason for a ”wrong choice” for ε might be that presentation order /bAk/ » /dAk/ was considered symmetrical to /dAk/ » /bAk/. Table 5.7 and 5.8 reveal this might not be the case; the main effect of B > D > B on γ_{cs} and n_c is not significant, but D > B > D is significant in both cases. Generating a parameter library with a larger range of ε values may solve this.

The suggestion to create a larger library lays bare an important practical drawback of this estimation method: It is computationally heavy. Finding the parameter estimates for the entire sample (including simulation and construction of the transition and emission matrices) can take more than 6 hours on a computer with a 2.3Ghz Intel i6 with 4GB of RAM. There is much room for improvement and efficiency of the method. One can imagine using a smaller pre-generated parameter space to get a gross estimate of initial parameter values which can be fed into a more common minimisation algorithm that can find the best parameters by dynamically simulating and calculating the likelihood of an observed series.

Another fact that may be seen as a shortcoming is that there was no model selection strategy. What if a linear model were better fitting? It should be possible to do some kind of model selection if the potential landscape could be transferred into a stochastic version like was done for the Cusp Catastrophe (Grasman et al., 2007; Wagenmakers, Molenaar, Grasman, Hartelman, & van der Maas, 2005) and the HKB model (Schöner, Haken, & Kelso, 1986). That is, however, beyond the scope of this chapter.

5.12 General Discussion

Well established and more recent phenomena in the empirical record pertaining to the relationship between speech perception and literacy development were re-evaluated and interpreted in terms of coupled potential functions representing the categorisation of some physical continuum by separate modalities or processes operating on different functional levels of the perception-action system. This notion of a coupling factor controlling the interaction of several processes or sources of information that is related to age is not new. Age often takes on the role of a collective variable summarising developmental maturation, academic history, life experience, etc. For instance, the cooperativity term used in the dynamic field model of infant perservative reaching, is such an age or maturation related term that governs whether the model behaves purely input driven (weak cooperativity) or whether processes on different timescales interact (high cooperativity) to define the overall model behaviour (see Thelen, Schöner, & Scheier, 2001).

Studying the temporal dynamics of speech sound categorisation as was done in Experiment 2 lays bare these differences and may warrant investigation into targeted interventions at a later age than is currently common. This explanation for the emergence of apparent discrete perceptual (or action) categories is very different from accounts that represent them somewhere in the nervous system or genes as a collection of perceptual codes of the boundaries of the categories. The results presented in this chapter seriously question both the discreteness and hardwired nature of perceptual categories:

Adding "noise" to the perceptual process can induce category switches, such that for one and the same value of a physical dimension, two different categories can be perceived (Experiment 1). This could be interpreted as within-category discrimination and evidence for a perceptual boundary.

The internal structure of perceptual categories, and the boundary between them measured along a physical dimension is not static, but dynamic. The structure inside a category does not start out as a smoothly defined space, which could be expected if millions of years of evolutionary effort were put into storing such boundary codes in genes and the nervous system. Rather, categories become smooth and well defined under influence of learning and development. This processes is subtle and continues over a wide age range (Experiment 2). Although previous studies acknowledged influence of learning such as literacy training on the perception of speech sounds (e.g. Noordenbos et al., 2012), a much shorter time range is assumed to cause disappearance of differences between average and dyslexic readers with respect to subtle speech perception differences such as within category discrimination. Studying the temporal dynamics of speech sound categorisation as was done in Experiment 2 lays bare these differences and may warrant investigation into targeted interventions at a later age than is currently common.

An interaction dominant model with associated coupling hypothesis is able to predict and explain observed speech sound categorisation behaviour by average and dyslexic readers. The model is based on a common class of potential functions frequently used to describe transitions between behavioural modes of systems under study in the physical, chemical and biological disciplines of science. The model predictions are made without invoking complex schemes of evolutionary information storage and retrieval. Its main assumption may be summarised as follows: Humans are able to divide any perceivable physical dimension into apparent static and discrete categories (when they are properly encouraged to do so by an experimenter).

What remains to be discussed is a larger theoretical framework for the interaction-dominant coupling hypothesis. To claim phonemes are not innate entities stored somewhere as information and even that they are not the higher order collective variable, by which humans perceive speech sounds, demands an alternative perspective.

5.12.1 The Ugly Duckling strikes back

Human speech is a complex signal, especially when taking into account the meaning it may convey as signal of communication. There are many parallels between human speech and bird vocalisations, including their use as a carrier of meaning: From the way chicks learn about local language variations by listening to adult birds, to the neurobiology of the bird song (Doupe & Kuhl, 1999). Especially interesting is the research into the phenomenon observed in ducklings who, after they just hatched, will only follow a mother duck if she produces the calls made by its native species (Gottlieb, 1991). Naturally, this has been interpreted as an innate ability, much like what has been discussed is assumed for human speech perception: The ability to perceive a signal of such complexity immediately after birth (hatching), must have been programmed by nature.

Careful studies show this is not the case at all. Duckling eggs were hatched in isolation, in order to rule out any auditory stimulation from its environment. This increased the probability of a duckling following the native call of another species somewhat, but not dramatically (Gottlieb, 1987). Further experiments with complete auditory deprivation yielded similar results. A remarkable experimental procedure was employed to provide the definitive answer. While the duckling was hatching inside the egg, its vocal cords were glued with temporary glue (devocalised) in order to prevent it from making any vocalisations while in the egg. This proved to be the causal factor for the emergence of the native call perceptual category: A duckling has to produce and perceive its own vocalisations while hatching in order to be able to recognise a call as its own (Dmitrieva & Gottlieb, 1994; Gottlieb, 1987). The necessary coupling is one of speech production and speech perception as alluded to in the description of the potential model presented in this chapter.

This phenomenon was studied in humans, yielding the famous result in which a specific text read by mothers during pregnancy was preferred by infants over a novel text (DeCasper & Spence, 1986). This result just concerns auditory experience (not production, which is more complicated in mammals) and is not a test of specific speech sounds like phonemes. A recent study examined on average 33-hour-old infants from Sweden and the USA on their perception of native English and Swedish vowels (Moon, Lagercrantz, & Kuhl, 2012). The infants had a preference for the non-native vowels or vowel prototypes indicating the native vowels were already familiar to them. These results were interpreted in the context of the perceptual-magnet theory (Iverson & Kuhl, 2000), because infants also preferred phonetic prototypes of non-native vowels over native vowel prototypes. The perceptual magnet theory states there are innate speech sound prototypes that represent the “initial state” of the speech code (Kuhl, 1991), which will be changed by exposure to speech sounds perceived in the environment.

The argument for innate exemplar speech sounds comes from the observation that perceptual magnet effects cannot be found in monkey (Kuhl, 1991). The little ducklings have shown however that what is innate, is more likely form and shape, or structure rather than anticipation on the outcomes of a perceptual process, or dynamical patterns. Because of the specifics of the body of the duckling, because of the specifics of the egg its mother produced, because of the laws of physics, the duckling develops the ability to recognise a complex auditory signal once it hatches. This is an example of nonlinear causation (Gottlieb, 1991), that gives rise to emergent properties and behavioural patterns in interaction-dominant systems. Nonlinear causation refers to the fact that stable and consistent forms and behaviour can emerge (or disappear) without apparent efficient causes in complex living systems (e.g. through reproduction, during development or in real-time interaction with others). Such dynamical patterns, like a speech sound category can only emerge due to a structure. The suggestion that is implied by innate speech sounds is that the body somehow contains within itself, those dynamical patterns as a structure as well. A mathematical equation does not contain all the patterns it can produce, the equation is the structure and the pattern resides in the interaction with time and other processes, but not as a separate entity. It seems to me that a change process like the evolution of species by natural selection can only efficiently provide a structure that allows dynamical patterns to emerge with a certain likelihood (e.g. development of a physical body configured such that stable “initial states” can emerge in a controlled environment like a womb or an egg). Storing patterns like speech sounds that are strongly emergent due to nonlinear causation as separate entities within the very structure from which they can also emerge dynamically seems… a highly improbable choice for a natural process.

5.12.2 Mismatched perseveration

In the introduction to this study I questioned the interpretation of the (sometimes decades old) empirical evidence of infant speech perception on which the allophone-coupling hypothesis is based. A recent physiological study of the brain provided evidence for allophone perception in children at-risk for dyslexia (Noordenbos, Segers, Serniclaes, Mitterer, & Verhoeven, 2012b), using the so-called mismatch negativity (MMN) paradigm, sometimes called error related negativity. This is a technique that is widely used in neuroscience, but when viewed from the perspective of potential model like the one presented in this chapter, its validity should be questioned. This can of course have grave consequences for the interpretation of results obtained using the technique. The paradigm entails the presentation of a stimulus that is repeated several times, after which a novel stimulus is presented. The novel stimulus is a mismatch with respect to the preceding sequence and elicits a typical brain potential that may be observed as a negative peak in the ERP, hence the term mismatch or error-related negativity. Speech perception may be studied in this paradigm by taking as the repeated and novel stimuli, sounds that are very similar, or close with respect to their ordering on some auditory continuum as discussed throughout this dissertation (Cheour-Luhtanen et al., 1995). Observation of a MMN related ERP would point towards a successful discrimination of the speech sounds; its absence would be interpreted as the brain not recognising a novel stimulus was presented.

The MMN response is measurable in infants even when they are asleep. Prospective studies have reported associations between reading ability at various stages of literacy development and infant speech discrimination measured in newborns to 17-month olds (Been, Van Leeuwen, & Van Herten, 2008; Guttorm, Leppänen, Hämäläinen, Eklund, & Lyytinen, 2011; Lyytinen et al., 2005; Molfese, 2000; Neuhoff et al., 2012; Pihko et al., 1999; Richardson, Leppänen, Leiwo, & Lyytinen, 2003; Tsao, Liu, & Kuhl, 2004; van Alphen et al., 2004). Interestingly, van Leeuwen et al. (2008) and other studies by this group (Been et al., 2008; Leeuwen, Been, Kuijpers, Zwarts, & Maassen, 2006; van Herten et al., 2008) used the same speech contrast (/bAk/, /dAk/) based on the same stimuli used throughout this dissertation (cf. Beinum, Schwippert, Been, Leeuwen, & Kuijpers, 2005). How can this large body of empirical evidence be interpreted in terms of the potential model? Most importantly, I question whether the paradigm measures a novelty response that can be interpreted as the perception of a mismatch or an error with respect to the previous stimuli. I suggest the paradigm is in fact a perseveration context that measures whether the novel stimulus is capable of perturbing the state the system settled in due to repetition. Consider a potential function with two possible states that represent perception of the repeated and the novel stimulus (Figure **5.16**). The potential landscape changes under influence of repetition of a stimulus such that the state gains stability at each repeated presentation n (e.g. Equation 5.2). At some critical value n_c the novel stimulus is presented, changing the potential landscape in such a way that the other state becomes a competing stable option. In Equation 5.2, the impact of the change in relative stability depends on the parameter ϵ.

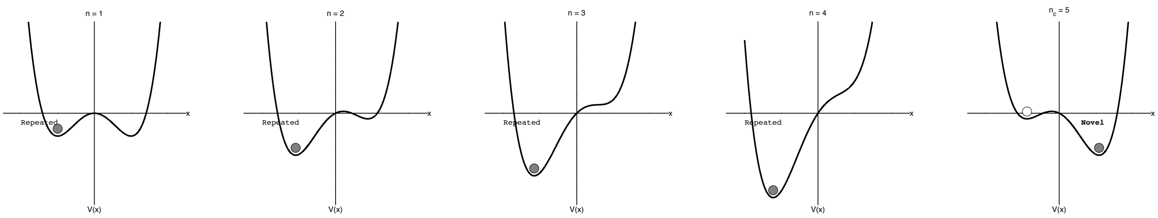


Figure 5.16 – Shown is a sketch of perception according to a potential function in a hypothetical mismatch negativity paradigm. There are four repetitions after which a novel stimulus is introduced. Depending on the parameters of the model, it is possible to simulate the nonlinear dynamics discussed in this Chapter: Critical jump (MMN response at n_c), Early jump (MMN response before n_c), Late jump (MMN response after n_c).

What is important is that the perseveration interpretation of the paradigm does not see the absence of an MMN wave as a failed discrimination of the repeated and novel stimuli. It means the Hysteresis or Enhanced Contrast phenomenon may have occurred instead of the critical jump phenomenon. What exactly shapes the potential landscape will not be very different from what was presented in this chapter: Coupling between continua of collective variables by which stimuli may be ordered, possibly by different modalities (γ_{cs}) and factors specific to the individual (ε) and the task (n_c, definition of V(x,y) and k(n)).

It is possible to re-analyse existing data and look for early or late jumps. This requires a different analysis approach than averaging to obtain ERPs however. In the study by Noordenbos et al. (2012b) two blocks of 400 stimuli were presented, each block had a .12 chance of occurrence of a novel stimulus (48 deviant vs. 352 standard stimuli). This means a maximum of 2x96 (the standard presented before the deviant is used for comparison) time series of about 700ms each were averaged, just a fraction of the total data recorded. On the trial level presentation is pseudorandom, ensuring the experiment starts with a reasonable number of standards and a number of standards always needs to follow presentation of a deviant. On average about 3-6 standard stimuli "surround" a deviant stimulus and this is quite similar to the number of stimuli used to represent a continuum in a standard categorical perception experiment (10 stimuli). An alternative analysis strategy based on a potential model could look for transitions in EEG recordings in a larger range around the deviant.

In order to look for transitions in time series data, Recurrence Quantification Analysis is the best tool to use (applications of RQA for this purpose, e.g. de Graag, Cox, Hasselman, Jansen, & de Weerth, 2012; Lichtwarck-Aschoff, Hasselman, Cox, Pepler, & Granic, 2012; Stephen, Dixon, & Isenhower, 2009). RQA has been adapted for EEG analysis and even specifically to detect negativity waves in a mismatch paradigm (analysis based on the Order Pattern Recurrence Plot, OPRP). Schinkel, Marwan and Kurths (2007) showed OPRP analysis was able to detect semantic mismatch negativity (N400) reliably based on ERPs consisting of 10 averaged time series. They were also able to detect the MMN in a single trial (Schinkel et al., 2007; Schinkel, Marwan, & Kurths, 2009). It seems there are no methodological objections to test the perseveration hypothesis.

There even is empirical evidence in support of this hypothesis that was discussed earlier in the context of Experiment 1 of this chapter. Tuller et al. (1994) induced category switches by repeatedly presenting the same stimulus at a point along the continuum where a category switch might be expected. They observed perception of the other category after 2-6 repetitions of the exact same stimulus! This is well within the range of sequences of standard stimuli presented before a deviant in the MMN paradigm. Moreover, in a speech perception MMN experiment the stimuli are usually not the end-points of the continuum, but are chosen more closely to the empirical phoneme boundary. Van Leeuwen et al. (2008) use stimulus number 3 and 6 as standard and deviant respectively, of the 10-step /bAk/-/dAk/ continuum used in Chapter 4 and Experiment 1 of this chapter (Been et al., 2008; van Herten et al., 2008; van Leeuwen et al., 2006 use the same stimuli). These studies report that discrimination difficulty was largest for dyslexic readers with stimuli between and including 3 and 6. This is indeed what was replicated in Experiment 1 (e.g. see Figure 5.6a and b), these stimuli were also the stimuli that were most prone to perturbation by noise. The conjecture is as follows: When encouraged to do so by a measurement context, dyslexic readers, but also preliterate children at risk for dyslexia, will reveal perception-action performance that may be characterised by loosely coupled processes operating at the collective level of emergent coordination (cf. Turvey, 2007; Van Orden, Kloos, & Wallot, 2009; Wijnants, Hasselman, Cox, Bosman, & Van Orden, 2012). In the measurement context of MMN, the observation of a novelty, or error response should be seen as the result of nonlinear interaction-dominant dynamics, for example dynamics described by (coupled) potential functions: Repetition of the same stimulus (the standard) that is only slightly different from its antagonist (the deviant) is known to cause a switch of states in speech perception experiments after 2-6 repetitions of the same stimulus. On average MMN speech perception studies repeat the standard (which is often indeed only slightly different from the deviant) 3 to 6 times. This habituation frequency is used in many studies and was established empirically for healthy subjects.

A low coupling strength can contribute to an increased likelihood of the occurrence of a jump between states because in general, the states related to the standard and the deviant will be less stable than states associated with a high coupling strength.

Absence of MMN may be due to hysteresis or enhanced contrast effects, not an inability to perceive the stimulus a deviant. Lower MMN amplitude may be indicative of less stable states in general (due to coupling strength). Analyses are available to test the current empirical record for early or late jumps with respect to the deviant and low or high amplitudes of waveforms in single trials.

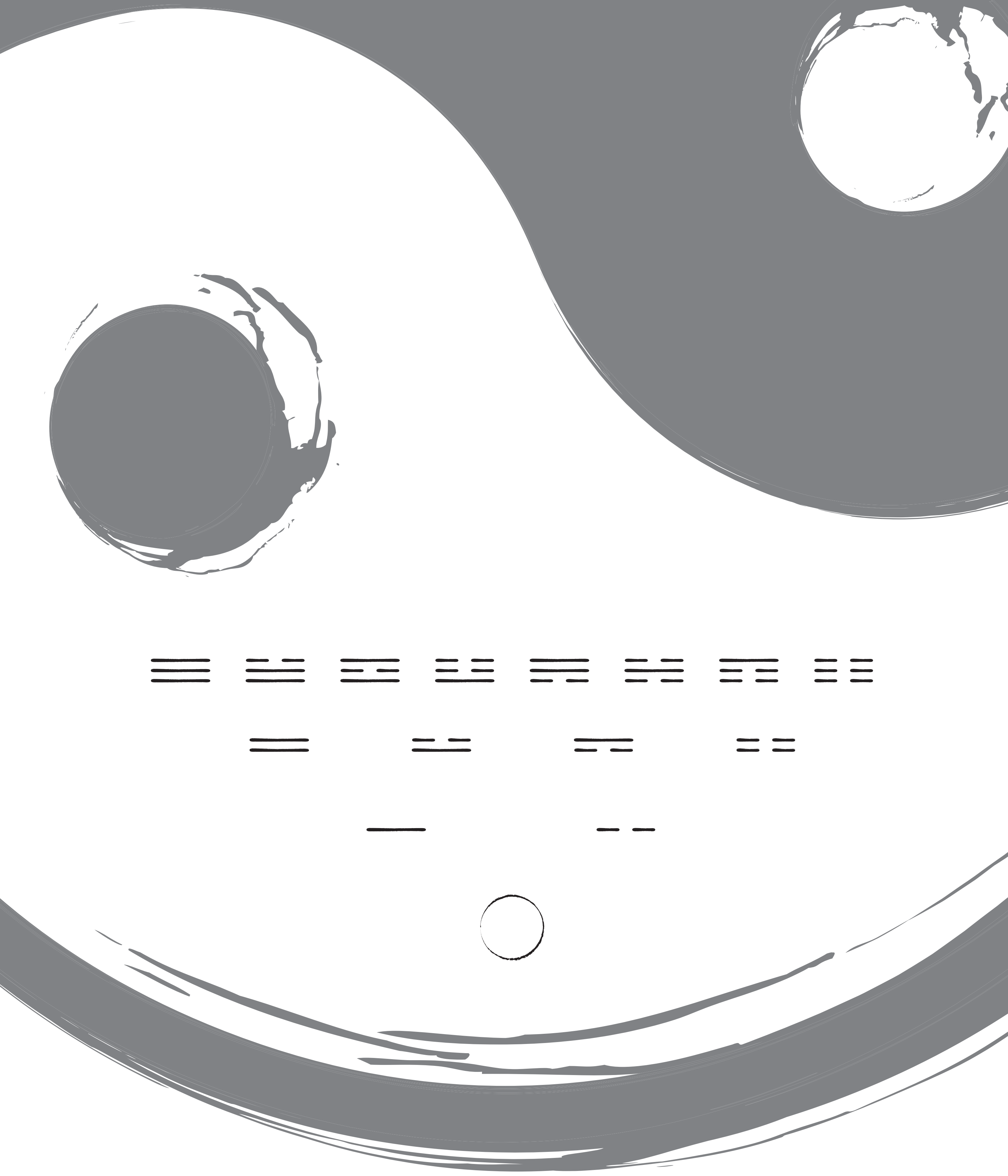
When an appropriate model has been deduced from existing data, model-based predictions about associations between variables like the ones presented in this chapter can be evaluated. Obvious experimental manipulations pertain to number of repetitions, timing of stimuli and coupling strength by comparing age across different (patient) populations.

The MMN study reported by Noordenbos et al. (2012b) also examined a standard and a deviant that belonged to the same phoneme category, but were opposites with respect to an allophone boundary. The results were as follows: "Stimuli from different phoneme categories elicited MMNs in both the control and at-risk children while the stimuli from the allophonic contrast elicited an MMN in only the children at risk for dyslexia." (Noordenbos et al., 2012b, p. 6). The at-risk children showed amplitude values of MMN for both types of stimuli that were quite similar in range (-2.5 to -1.5) whereas the control children had a much larger range (0.2 to 4.1) in amplitude. All the amplitudes around zero were observed with the allophonic contrast. The authors do not explain this result, but it is readily interpretable in terms of a weak coupling where the relative stability of states is more equal than with strong coupling that can cause disappearance of states altogether while boosting others.

The authors interpret the results as follows: "[…] the phonological representations needed for effective grapheme–phoneme mapping are not properly acquired by children at risk for dyslexia. It can be concluded that the phonological representations of the control children in the present study were more accurate, as reflected by their significantly better performance on the skills relevant for the development of reading, leading to more rapid selection of the relevant cues in the auditory signal." This is again a circular argument, much like the one discussed in the introduction. The phonological representations of average readers are more accurate, we know this because they perform better on reading and associated tasks, and this causes them to be better at auditory processing, which is why their phonological representations are more accurate.

To conclude, the time to peak amplitude of MMN was significantly slower for children at risk for dyslexia and this was interpreted as slower processing of auditory stimuli due to less accurate phonological representations. The question is the same as for the amplitude, why was this latency equally slow for the phoneme and the allophone deviant? If children at-risk for dyslexia have less accurate phoneme representations, they must have more accurate allophone representations, why do these natural categories not lead to faster peak amplitudes whereas the difficult and learned phoneme categories do in control subjects? Again the answer must be an unstable potential landscape, as was evidenced by the response time dispersion in Chapter 2 and 3.

Summarising, the current empirical record, theory evaluation, experimental results, and model simulation can together provide an alternative perspective on component-dominant theories of reading and dyslexia. Such theories often appear to take shortcuts around the big questions: How did the phoneme representations get there? How can a dynamical pattern be part of a structure that generates those dynamical patterns? Where does a cause begin and an effect end? There are many directions for future studies of the framework and the model presented here to explore. The most interesting suggestion is perhaps a re-analysis of the MMN paradigm. If the conjecture were validated, this would have substantial consequences for an entire field of science. But that is beyond the scope of this dissertation…



The Limitless produces the Delimited, and this is the **Supreme Unity**
[or: the **Absolute**]

The Supreme Unity produces the two **Forms**, named Yin and Yang
The 2 Forms produce four **Phenomena**,
named Lesser Yin, Great Yin, Lesser Yang and Great Yang.
The 4 Phenomena give rise to the 8 **Fundamental principles of reality,** [or trigrams],
8 eights are 64 hexagrams."

- Translated from ancient Chinese writing and ascribed to Fu Xi (± 2953-2838 B.C)
in The Book of Changes (Legge, 1963)

From left to right the 8 principles of reality are: Heaven, Lake, Fire, Thunder, Mountain, Water, Wind and Heaven. Fu Xi was the first of the 3 Sage Emperors of ancient China: A cultural hero, with mythological and god-like abilities (compare to Carolus Magnus, but roughly 4000 years earlier). He was said to be the survivor of a great flood together with his sister. With permission of the Emperor of Heaven they tried to re-populate the Earth with humans. To speed up the process and backed by divine powers they started to **create humans from clay.** He is thought to have created many things among which are the Ba Gua trigrams (Cammann, 1991).

Apocrypha

Cammann, S. (1991). Chinese Hexagrams, Trigrams, and the Binary System. Proceedings of the American Philosophical Society, 135(4), 576-589.

Legge, J. (1963). The I Ching: The Book of Changes (J. Legge, Trans. 2 ed.): Dover Publications Inc, New York, NY.

Notes | **Chapter 6: General Regularities & FirstPrinciples**

- A substantial part of this chapter will appear in:

  Hasselman, F. (pending revisions). Radical Embodied Computation: The Brain as a Self-Tuning, Self-Affine Resonator. Research Topic: Radical Embodied Cognitive Neuroscience. FrontiersIn Human Neuroscience.

- This chapter may be cited as:

  Hasselman, Fred (2014): A General Discussion of Principles The Role of Internal Representations in Ætiologies of Developmental Dyslexia. In *Beyond the boundary. An analysis of verisimilitude and causal ontology of scientific claims: Ætiologies of developmental dyslexia as a case in point*. fig**share**. http://dx.doi.org/10.6084/m9.figshare.1086257

- The supplemental materials to this chapter may be cited as:

  Hasselman, Fred. "Beyond The Boundary - 6th Chapter: Supplemental Materials." Open Science Framework.

- For Guy, a true master of the improbability drive.

Chapter 6

A General Discussion of Principles

The Role of Internal Representations in Ætiologies of Developmental Dyslexia

> "So what if I can't spell Armageddon, It's not like… the End of the World"
>
> —Stewart Francis

The general topic of this dissertation has been the evaluation of claims of scientific explanation about a domain in reality, in the presence of a great diversity of competing theories. As a point in case and due to the large amount of high quality data from various different disciplines and experimental paradigms, the evaluation concerned theorising about the causes of developmental dyslexia. More specifically, the component dominant ontology of phoneme representations and their presumed anomalous composition that is used to explain impaired reading was compared to an interaction dominant ontology of coupled dynamics of collective variables from which a landscape of potential perceptual events emerged. The main incentive for such a comparison is the current state of paralysis that characterises the empirical record: A plethora of ideas, hypotheses, and theories about what causes the impaired reading and spelling ability of a small and specific subgroup of developing children are posited, without there being a clear "winner". There is no theoretical account that can formally be shown to be more "truth-like" than others. My claim throughout this book has been that this is a general problem of the empirical social sciences and that one of its causes is that the implicit causal ontology escapes falsification because it is accepted as a part of reality. Theory specification does not evolve beyond Stage B of Box **1.4**. In Stage C the functional form of relationships between constructs is subject to empirical inquiry. Very rarely does the functional form move beyond linear.[1] This opinion I share with many other scientists (see e.g., Stepp, Chemero, & Turvey, 2011), but is best expressed by Stephen and Van Orden (2012). They respond to the "we want components" critique (see section 1.5.4 of Chapter 1) of the Complex Systems Approach:

> "Although we are told that the cognitive theorists have considered the interactions among mind, body, and environment, at the end of the day the only factors that matter are latent cognitive factors (Wagenmakers et al., this volume).

In other words, several commentators have decided what will count as a viable cognitive explanation. If we or anyone else refuses to play by their rules, they threaten to take their ball and go home, which might be OK if there were a strong tradition of successful empirical cognitive science, arriving at reliable explanations that are widely recognized inside and outside of cognitive science. Were that the case it would be prudent to remain skeptical of "revolutions" or "paradigm shifts." But over half a century, empirical cognitive science has had its own difficulties winning over hearts

[1] Note that the Fourier transform can be seen as a regression analysis with parameters of sine and cosine functions as the linear predictor. A higher order polynomial regression (e.g., y = x + x2 + x3) can also model a nonlinear shape, but both represent a linear functional form. Nonlinearity implies multiplicative interactions.

and minds, even without our meddling" (Stephen & Van Orden, 2012, p. 97) This final chapter is a discussion about the principles behind the two causal ontologies that appear to some as different ball games altogether. It will be a retrospect to what was learned empirically and a prospect to where future developments may head.

6.1 Results revisited

The conjectures put forward in the first chapter were: 1) A theoretical account that could replace current theories of developmental dyslexia should provide an explanation of the apparent lack of authority of empirical evidence to decide between the veracity of competing theoretical claims; 2) It is the kind of theorising practised in the soft empirical sciences that is responsible for the weak knowledge base (also see the Preface). To justify these conjectures I provided meta-theoretical, statistical and historical arguments, some of which will return in later sections of this chapter. Most of the presented arguments continue to play an important role in contemporary discussions of scientific practice in empirical social science (e.g., Asendorpf et al., 2013; Lebel et al., 2013; Open Science Collaboration, 2012). I concluded Chapter 1 by announcing empirical tests of the interaction dominant ontology that would attempt to move beyond 'stage B' and in this section I briefly discuss the extent to which the reported studies succeeded to achieve that goal.

Chapter 2 – Context relativity of ordering relationships and prospective prediction:
Predicting Reading Performance from Pre-Literate Speech Perception in Children at-risk for Dyslexia

The second was a cautious attempt to readjust the quest for the ultimate efficient cause towards an effort to understand what it is that makes multiple ætiologies appear as a scientifically plausible causal factor associated with developmental dyslexia. The design and analysis strategy of the empirical study allow for a conclusion beyond the effects of factors that obfuscate verisimilar associations between variables identified in Chapter 1. The conclusions were based on: 1) Prospective prediction, 2) coherent explanation and control of the appearance and disappearance of associations. The purpose of this chapter was to provide a different perspective on the causes of impaired performance as emerging in specific contexts from the interactions between hypothesised components, rather than originating from (deficient) components, in other words, the exquisite context sensitivity associated with interaction dominant dynamics (Riley, Shockley, & Van Orden, 2012). Context relativity was examined for the stability of speech perception as expressed by the dispersion of response times rather than the accuracy of the performance. In addition, the role of context on speech perception was investigated by measuring speech perception in different participant groups (familial risk vs. no risk) using different tasks (identification vs. discrimination) and different acoustic manipulations of the speech stimuli. The results showed a clear relation between Grade 1 reading performance and stability of kindergarten speech perception: Unstable speech perception was associated with lower levels of later reading performance. The accuracy of performance did not reveal such associations.

Chapter 3 - Principled simulation of context relativity:
When opposites attract, repel and deceive: Using Recurrent Neural Computation to Model Multistable States.

In order to understand the dynamics between speed and accuracy of a classification performance in stable and unstable systems, Chapter 3 explored the attractor dynamics of a recurrent neural network (Hopfield, 1982) between two opposing basins of attraction. These attractors represent the classification of a stimulus as either /bAk/ or /dAk/ based on exemplar values along two dimensions that were imagined to represent the features of the speech signal (F2 salience and F2 rate of frequency change) identified as important by the auditory temporal processing deficit hypothesis (cf., Tallal, 2004). In principle (and in the light of the results of Chapter 4), these dimensions can be assigned any meaning. Classification accuracy was examined by presenting the network with stimuli that vary along the two dimensions. System instability was introduced as (but not limited to) the presence of a third, weak basin of attraction that would never return a stable solution (i.e., no errors would be made), but still disrupted the energy landscape sufficiently to increase the duration with which stable solutions would occur. The latter was shown to be associated to the larger dispersion in response times observed in some of the measurement contexts of Chapter 2.

Recently more sophisticated methods have been developed to analyse response time dispersion using distribution analyses (Holden, Van Orden, & Turvey, 2009; van Rooij, Nash, Rajaraman, & Holden, 2013) and fluctuation or fractal analyses (Holden, 2005). Overall the results obtained in these studies confirm that dispersion of measurements can characterise stable responses from overly random or overly constrained responses. These techniques are now widely used to evidence interaction-dominant dynamics in human physiology and performance (Hasselman, 2013; Ihlen, 2012; Ihlen & Vereijken, 2010; Kello, Beltz, Holden, & Van Orden, 2007; Van Orden, Kloos, Wallot, & Orden, 2009; Wijnants, Cox, Hasselman, Bosman, & Van Orden, 2012; Wijnants, Hasselman, Cox, Bosman, & Van Orden, 2012).

Chapter 4 - Strong Inference:
Classifying acoustic signals into phoneme categories

In this chapter several features of speech stimuli that lie on a /bAk/-/dAk/ continuum were tested on their ability to enable a simple classifier (Quadratic Discriminant Analysis) to reproduce the observed classification performance of average and dyslexic readers. It was attempted to create a test of theoretical claims under conditions of strong inference, which requires the positing of deeper-lying entities (Hasselman, 2013; Platt, 1964). The 'classical' or represented features were based on component process accounts of developmental dyslexia such as the supposed deficit in Envelope Rise Time detection and the deficit in the detection of rapid changes in the distribution of energy in the frequency spectrum (formant transitions). Remarkably, studies examining these temporal processing deficit hypotheses do not employ measures that quantify the temporal dynamics of stimuli. It was shown that measures based on quantification of the dynamics of complex, interaction-dominant systems such recurrence analysis and multi-fractal scaling enabled QDA to classify stimuli almost exactly the same way as actual dyslexic and average reading participants did. It thus seems unlikely that participants used the measures, associated theoretically with faulty representations or component processes, when classifying the stimuli. Results were interpreted to support of the *Complexity Matching Hypothesis* of perception and action.

Chapter 5 - Principled simulation of posited entities and strong inference:
Beyond the Static Phoneme Boundary: The Nonlinear Dynamics of Emerging Literacy

In Chapter 4 it was shown that higher order variables quantifying complex temporal patterns in the speech signal were most likely the characteristic properties of speech sounds that dyslexic and average readers use to categorize different variations of /bAk/ or /dAk/ as such. I suggested that speech perception should not be considered a matter of analysing frequencies or amplitude patterns in order to match those patterns to exemplars of speech sounds stored in the brain, perhaps installed there by evolution. Chapter 5 promoted the idea that the existence of separate speech sounds, such as phonemes, are actually due to the use of an alphabetic script. Apparently, Chinese adults proficient only in the character script do not perceive any separate speech sounds at all. This prompted a critical analysis of the decade old empirical 'facts' about speech perception and their use by contemporary theories of dyslexia. The interpretation of the facts by accounts such as the perceptual magnet theory and the allophonic mode of speech perception hypothesis seemed somewhat biased towards the respective theoretical claims. Both accounts claim innate phoneme prototypes exist (either as boundaries or as perceptual categories) in the nervous system, but this is not an explanation of the claim. However, the general theoretical idea put forward by these accounts about the causes of developmental dyslexia conjectured as a deficient coupling between different sources of perceptual information was taken as a sound suggestion to depart from.

In line with the prescriptions of Chapter 1, a theory of principles was constructed and a model was suggested based on interaction dominant dynamics governing speech perception of dyslexic and average readers. The model allowed for risky predictions of the kind that can be considered a "damn strange coincidence" had they been observed absent the theory. The differences between dyslexic and average readers, for instance, could be defined as specific tests of the coupling hypothesis: Coupling strength between two dimensions representing different collective variables along which different speech sounds can be ordered. The model was conjectured to be able to explain a rich variety of nonlinear phenomena such as hysteresis and enhanced contrast perception.

Results showed that it is able to provide a more parsimonious explanation of the better within category perception by dyslexic readers that prompted the allophonic mode of speech perception hypothesis (Serniclaes, 2006). Two empirical speech perception studies in which a large group of children, varying in age and reading ability participated, reveal the nonlinear dynamics of speech perception as predicted by the model. That is, the empirical results can be understood as resulting from different parameter settings of the model (by considering each trial in the experiment the result of the emissions of a Hidden Markov Model whose probability matrices are based on the stability of the potential landscape). An analysis of the fitted parameters should be considered a corroboration of the predictions by the interaction-dominant coupling hypothesis.

The principles on which the models of Chapters 3 and 5 are based remain a topic of study; especially the possibilities of inferring model parameters from empirical data. Recent studies explored fitting the parameters of the 1-dimensional potential model in the context of hysteresis in decision making (van Rooij, Favela, Malone, & Richardson, 2013; van Rooij & Favela, 2010; van Rooij & Van Orden, 2011). There have been advances in modelling and fitting negative hysteresis or enhanced contrast (Lopresti-Goodman, Turvey, & Frank, 2013) Studies of two dimensional neural networks such as the one in Chapter 3 reveal there may exists a topological equivalence between the energy field of such networks and the potential field of the two-dimensional potential model of Chapter 5 (Chartier, Renaud, & Boukadoum, 2008). Future directions could explore the connections between these topologies of deterministic dynamics and those of critical state neural networks, such as the critical branching network (Kello, 2013), or the Potts model with hidden states (Tanaka, Tamura, & Kawashima, 2011).

6.2 Definitions revisited

> "Here I have completed this bird's-eye survey of the principles that govern the self-organizing system. I hope I have given justification for my belief that these principles, based on the logic of mechanism and on information theory, are now essentially complete, in the sense that there is now no area that is grossly mysterious." (Ashby, 1947)

In what way do the results evidence an interaction dominant ontology or question the viability and veracity of the phoneme representation ontology? Has a theory of principles been tested and can the conjectures of the first chapter be raised to postulates? To try and answer these questions, I will reconsider and elaborate on some definitions provided earlier, as well as strengthen some claims with additional evidence.

6.2.1 "How would we recognize it?"

Many theories claiming explanatory authority on the causes of developmental dyslexia are based on an idea, a blueprint of the cognitive mechanism giving rise to human reading and spelling ability: It must be some process that transforms letters into sounds via their abstract representation in the nervous system. The representations are in this case the components, or entities of the ontology, because the theory posits them to exist as constituents of reality. That is what scientific theories are supposed to do of course, positing of deeper-lying entities. Curiously, what constitutes a representation is not clearly defined or operationalized in fields of psychology such as behavioural and cognitive science (Haselager, de Groot, & van Rappard, 2003). Haselager et al., tried to answer the question posed by Cliff and Noble (1997, p. 1170): "if evolution did produce a design that used internal representations, how would we recognize it?" Of course, the same can be asked of the alternative ontology: "If evolution did produce a design that used coupled dynamics across many time and spatial scales, how would we recognize it?" Note that the latter definition does not exclude the possibility of such systems' internal states representing external physical states. There are several crucial differences though, but before I discuss them, consider the definition of representation provided by Haselager et al. (2003):

> "We consider the issue of representation from a realist perspective. That is, "representation" refers to an identifiable physical state within a system that stands in for another (internal or external) state and that as such plays a causal role in (or is used by) the system generating its behavior." (Haselager et al., 2003, p. 6)

The specific claim related to dyslexia is often that representations of speech sounds that stand in for the external sound/pressure waves generated by human vocal chords, must be a degenerate or anomalous version of the actual external physical state in order to explain impaired acquisition of reading ability. As shown in Chapter 1 it is a logical fallacy to infer that a structure exists based on an observed phenomenon and this also applies if a modifier like "impaired" is attached to the phenomenon. Such inferences are driven by a perceived isomorphism ("is similar to") between the posited representation ontology and the observed behaviour. Detecting an isomorphism is not a valid strategy for evidencing representations in a physical system and Haselager et al. (2003) conclude there does not exists an operational definition of representation that is adequate to make plausible their existence empirically.

I do believe questioning the representation ontology empirically is possible as long if it is done in the context of strong inference (cf., Platt, 1964). Chapter 4 is an example of such an inquiry that can be characterised as a test of the different properties attributed to internal representations of external states (stimuli) by different theories. An explicit formal or operationalized definition is not necessary if these features are directly compared to features that are based on a different ontology. In the light of the results presented in Chapter 4, it is quite clear that the measures laminarity and determinism, quantifying the dynamics in the reconstructed phase space of an auditory stimulus, as well as the multifractal spectrum, are of a different nature than frequencies and amplitudes of the physical signal. This does of course not prevent a promiscuous theorist like investigator James to suggest it must be these dynamical properties, determinism and laminarity, these values of singularity spectra, that are represented internally. To show how odd it is to suggest this, it would be the identical to a theorist claiming the F value of his significant General Linear Model was represented internally.

A suggestion more viable, would be that one or more system parameters are set in such a way that a perturbation of the system by the physical signals that can be characterised by these dynamics, cause it to settle in the observed behavioural mode. The parameter settings are nothing more than modelling abstractions, but they do "represent", in the model, a composition of the system. This kind of representation reflects dynamical variables (Spencer & Schöner, 2003), it suggests that instead of a single 'true parameter', a larger parameter space can be explored by a system (Jacobs &

Michaels, 2007; Michaels, Arzamarski, Isenhower, & Jacobs, 2008) and can be understood in terms of the physical concept of a synergy (Akhromeeva & Malinetskii, 2009; Haken, Kelso, & Bunz, 1985; Kugler & Shaw, 1990; Schöner, Jiang, & Kelso, 1990; Turvey, 2007).

How would we recognise a synergy? Synergies are large-scale organisations of system components that solve the degrees of freedom problem: The degrees of freedom available to produce speech sounds are fewer than those that would be possible by permutations of speech-sound components (Galantucci, Fowler, & Turvey, 2006; van Lieshout, Kent, & Lieshout, 2004). The fact that components of a system are physically coupled, by the neuromuscular structure of the speech apparatus, the respiratory system and enslaved to the laws of thermodynamics that govern the behaviour of gasses, the degrees of freedom available to the system to produce sounds are constrained. Synergetic coordination of behaviour is observed in systems that are characterised by ultrafast action, ultrafast cognition, exquisite context sensitivity and scale-free variation (examples of synergetic control in cognition and behaviour are given in Chapter 5 and Riley et al., 2012; Wallot & Van Orden, 2012). Specific predictions can be formulated based on synergetic control context sensitivity and scale-free variation that are divergent from causal component predictions (Diniz et al., 2011).

In Chapter 5 the dynamical nature of such collective parameter settings that represent system compositions were evidenced at the time scale of the experiment (settings of the ε parameter associated to hysteresis, enhanced contrast, or critical boundary) and at the scale of biological development (the coupling strength parameter γ_{CS}). Chapter 2 evidenced the exquisite context sensitivity. A recent study on reading fluency in developmental dyslexia revealed strong evidence of scale-free variation in response latencies of word-naming that was correlated with reading performance measured on standard reading tasks, but only in dyslexic readers (Wijnants, Hasselman, et al., 2012). Then, why do so many scientists observe components when they study behaviour and cognition? Wagenmakers et al. (2012), critical of the complex systems approach, noted that studies have shown that standard models of cognition could sometimes be seen as special cases of nonlinear, models with many interacting components. They conjectured that the tension between old science and the complex systems perspective might only be apparent, as the latter justifies the former. This conjecture is not new however and has been known to psychological science since the 1940s through the work of W. Ross Ashby published in the Journal of General Psychology (Ashby, 1945, 1947). It covers exactly the problems with the complex systems approach reported by Wagenmakers et al. (2012) in Chapter 1. Ashby provides two arguments for the proposition that mechanisms and components may be observed in complex systems that are actually self-organising holistic structures.

The first is what I have been arguing for throughout this book; organization is in the eye of the beholder or the measurement context for that matter. Two researchers observing the same beehive may conclude differently about its organization. One sees the hive as the result of the interactions of 50.000 bee-parts and concludes a high level of organization, while another who observes whole states, such as activity and dormancy, may observe no organization at all (Ashby, 1962, p. 259). The second argument is about the independence of dynamics and organization or composition. A system, that appears an irreducible whole of interacting parts, can always be shown to consist of separate parts by suggesting what a part might look like and assuming a similar (isomorphic) part exists. If both parts are isomorphic to the whole, the parts will indeed be observed. Ashby provides a formal example of this phenomenon (not provided here) and concludes:

> "Thus, subject only to certain requirements (e.g. that equilibria map into equilibria), any dynamic system can be made to show a variety of arbitrarily assigned "parts", simply by a change in the observer's viewpoint." (Ashby, 1947).

That about covers conjecture 1 of Chapter 1: Even if a complex interaction dominant ontology is postulated, components and component dynamics (mechanisms) can be identified, quite arbitrarily.

It is important to note that this explanation of component mechanisms in terms of a feature of complex systems is supported by the strongest criticasters of the Complex Systems Approach to cognitive science (Wagenmakers et al., 2012).

I expand on this consensus explanation by suggesting the following: The default ontologies for many theories in psychological science today arise due to the adoption of the experimental methods of classical physics intended to study classical systems, Newton's Curse. A component configuration is postulated based on a singular behavioural phenomenon and this is the observer's viewpoint. By experimentation, parts that are isomorphic to the postulated components can be identified as "effects" of an independent variable, thus evidencing the component as a part of the internal structure of the system by means of association. The only way to be certain that a structure in reality was evidenced is by further theory specification, by prediction of additional observational constraints beyond the mere association. The ability to observe components at any level of analysis may be a feature of interaction-dominant systems rather than an error of the experimenter (that is, the error is in the interpretation, not in the data). Ashby explicitly mentions the self-organising systems he has in mind are not restricted to classical mechanical systems governed by Newtonian dynamics (Ashby, 1945, p. 14).

When Ashby's theoretical work is combined with that of Lashley, who spent a lifetime searching for the engram, the representation of memories by a physiological trace in the nervous system, it seems that the stage was set for a cognitive neuroscience based on self-organisation in complex systems in the early 1950s. It is striking how well Lashley's conclusions about representation of external states by the nervous system fit with the interaction dominant perspective, especially considering the fact that in-vivo brain-imaging was not available at the time. Here I quote Lashley's conclusions of an entire scientific career spent searching for empirical evidence of internal representation:

1. It seems certain that the theory of well-defined conditioned reflex paths from sense organ via association areas to the motor cortex is false. The motor areas are not necessary for the retention of sensory-motor habits or even of skilled manipulative patterns. [...]

2. It is not possible to demonstrate the isolated localization of a memory trace anywhere within the nervous system. Limited regions may be essential for learning or retention of a particular activity, but within such regions the parts are functionally equivalent. The engram is represented throughout the region. [...]

3. The so-called associative areas are not storehouses for specific memories. [...] The defects which occur after their destruction are not amnesias but difficulties in the performance of tasks which involve abstraction and generalization, or conflict of purposes. [...]

4. The trace of any activity is not an isolated connection between sensory and motor elements. It is tied in with the whole complex of spatial and temporal axes of nervous activity which forms a constant substratum of behaviour. Each association is oriented with respect to space and time. [...]

5. Consideration of the numerical relations of sensory and other cells in the brain makes it certain, I believe, that all of the cells of the brain must be in almost constant activity, either firing or actively inhibited. There is no great excess of cells which can be reserved as the seat of special memories. [...]

6. The equivalence of different regions of the cortex for retention of memories points to multiple representation. Somehow, equivalent traces are established throughout the functional area. Analysis of the sensory and motor aspects of habits shows that they are reducible only to relations among components which have no constant position with respect to structural elements."

(Lashley, 1950, pp. 26–27)

Conclusion number 6 has been repeated throughout this book in one form or another. The advancement in fundamental scientific knowledge after 60+ years of research on understanding the role of the nervous system in coordinating adaptive behaviour of living systems has been... disappointing, to say the least. On a more positive note against the background of my plea for formal theory evaluation in the social sciences, Lashley arrived at these conclusions by showing a rigorous ontological commitment. That is, certain propositions were certainly false in his opinion, because he actively explored the explanatory boundaries of the theoretical construct of internal representation of specific states. In doing so, he realised that the construct breaks down as an explanatory vehicle when it is confronted with a crash-test against reality.

To summarise, the interaction-dominant ontology does posit components to exist, it does not attribute any causal primacy to those components and it excludes internal representations of external physical states to be the structural components of the system. Observed modes of behaviour are due to the interactions between components conditional on the composition of those structural components of the system. In other words, the structural composition of the system is permissive of the modes of behaviour rather than causative. The objects of scientific inquiry based on this ontology are always collective variables, 'complexes', that can be conceived of as synergies represented in dynamic systems models by a configuration of parameter values. This type of causal ontology is not new; it originates from disequilibrium physics and chemistry, but has been successfully applied by the sciences that study biological systems or agent-environment systems whose behaviour is characterised as originating from many processes interacting on many different temporal and spatial scales (Ashby, 1947; Bak, Tang, & Wiesenfeld, 1987; Bunge, 1977, 2000; Heylighen, 1989; Oltvai & Barabási, 2002; Tononi, Edelman, & Sporns, 1998; Van Orden et al., 2009).

6.2.2 Caught in a Nomological Net: What constitutes a disease entity?

Understanding, in general, how sensible scientists may infer component ontology does of course not necessarily imply that such a mechanism must underlie the paralysis of the empirical record in studies on the causes of developmental dyslexia. In Chapter 1, I suggested to theoretically consider developmental dyslexia as a state, a complex INUS condition. I did not discuss, in terms of metatheory, how a specific ætiology may be defined, a "disease entity" in a nomological net. Some very insightful suggestions are provided by Meehl (1977) who concludes an article entitled "Specific Etiology and Other Forms of Strong Influence: Some Quantitative Meanings" with the following four sentences:

1. In summary, a disease entity is initially defined implicitly or contextually, loosely and not strictly "operationally," by the researcher- clinician's presenting a cluster of symptoms, complaints, and signs that covary over the population of patients (and usually over time in the individual patient).

2. This observational finding suggests the conjecture that the cluster, usually called the "syndrome," exhibits its statistical togetherness because of a causal source shared in common by the several indicators. At this early stage of knowledge, the disease entity is an "open concept," in the philosopher's sense.

3. The postulated causal relationships between the conjectured etiology and each of the fallible indicators provide at one and the same time a set of synthetic factual claims about the way the world is and, paradoxically, the implicit contextual definition of the entity.

4. Three aspects of the openness of these concepts are discussed, to wit, (1) the extensibility of the indicator list, (2) the probabilistic rather than strictly nomological correlation between the indicators and the disease entity, and (3) "Orphan Annie's eyes,"[2] that is, the to-be-researched inner nature of the entity (as we expect it to be reduced to lower levels in the pyramid of sciences)."

[2] "Orphan Annie's eyes" refers to a cartoon character Meehl (1977) used to explain this characteristic of open concepts.

(Meehl, 1977, p. 51 numbers and paragraph format added)

These four sentences constitute valuable scientific observations to absorb for the researcher-clinician, scientist-practitioner or anyone otherwise inclined to proclaim to have evidenced a disease, pathology, disorder or impairment of sorts. As for developmental dyslexia, the early stage of knowledge, as an "open concept" (Meehl, 1986, 2001) is certainly at hand and perhaps surpassed in that several causal sources have been identified that provide a "statistical togetherness" and at the same time provide a contextual definition for the disease entity (this evidencing and defining at the same time is due to the "Ramsified upward seepage" discussed in Chapter 1 and the Preface). This is one of the reasons why I opted out of the ætiological definition of developmental dyslexia in Chapter 1 and remained close to an operational definition: Everything else being equal, these children have difficulty acquiring proficient spelling and reading ability.

The three characteristics mentioned in the fourth sentence fully apply to the state of conjectured ætiologies: The indicator lists can be expanded, shortened, or changed; most indicators are probabilistic associations to the disease entity and if they are nomological deductions, they are inspired by what the "true" nature of the disease entity is imagined to be at the biological (physical) level. This inner nature is reflected in Figure 1.1 of Chapter 1, where the factors at the biological level serve as the ultimate efficient causes. This resonates with the concept of the interpretation fallacy and the identification of mere isomorphism as discussed in the previous paragraph. The observed associations imagined as "true" ontological causes, guide empirical inquiries. This bias may have occurred when I presented the graph networks of terms used in abstracts of articles in Chapter 1 (section 1.2). I categorised the terms into common factors used in the literature on developmental dyslexia in order to reduce the size of the graphs. In the process I may have excluded important terms or reduced variability to the extent that the network structure is no longer a good representation of the nomological net as it may be identified in a scientific record. Moreover, this great reduction in variability precluded me from conducting an analysis that could have directly tested the claim of a weak knowledge system. As has been known for some time, networks with scale-free (small-world) or small world structure are the common natural networks found in the physical, biological and social sciences (Barabási & Bonabeau, 2003; Bullmore & Sporns, 2009; Strogatz, 2001; Watts & Strogatz, 1998). This is attributed to optimal information flow (e.g., the "six-degrees of separation" in the global social network) and resilience to node failures or disconnections between nodes in the network (Stam, 2010).

Figure **6.1** is an adaptation of two figures originally published in the famous article by Watts and Strogatz (1998): "Collective dynamics of 'small-world' networks". In this article they described networks that had nodes whose local average degree was much larger than that of randomly connected networks of the same size (n) and neighbourhood (k), whereas the path lengths were similar or larger to such randomly connected networks. Besides ubiquity in nature (the brain and central nervous system are 'small-world' network structures see Bullmore & Sporns, 2009; Rubinov & Sporns, 2010; Stam, 2010), this composition of components has been shown to be an optimal information carrying structure. For example, the spread of infectious disease occurs fastest in populations organised according to small-world structure (Watts & Strogatz, 1998). Should an efficient scientific record be structured according to a small-world architecture, forming a strong knowledge base? If the goal of scientific explanation in the daily work of a researcher is exposing causal laws that should ultimately lead to unification as described in Chapter 1, it seems plausible that such a network of connected facts should ultimately collapse into a structure in which several concepts receive many connections through which most of the other nodes in the network can be derived. Also, if one sub network representing a theoretical whole should be separated from the larger network, pruned because it is falsified, this should not make the entire structure collapse. The parts that have not been pruned by falsification should still be reachable and this is one of the properties of small-world networks (Barabási & Bonabeau, 2003; Stam, 2010).

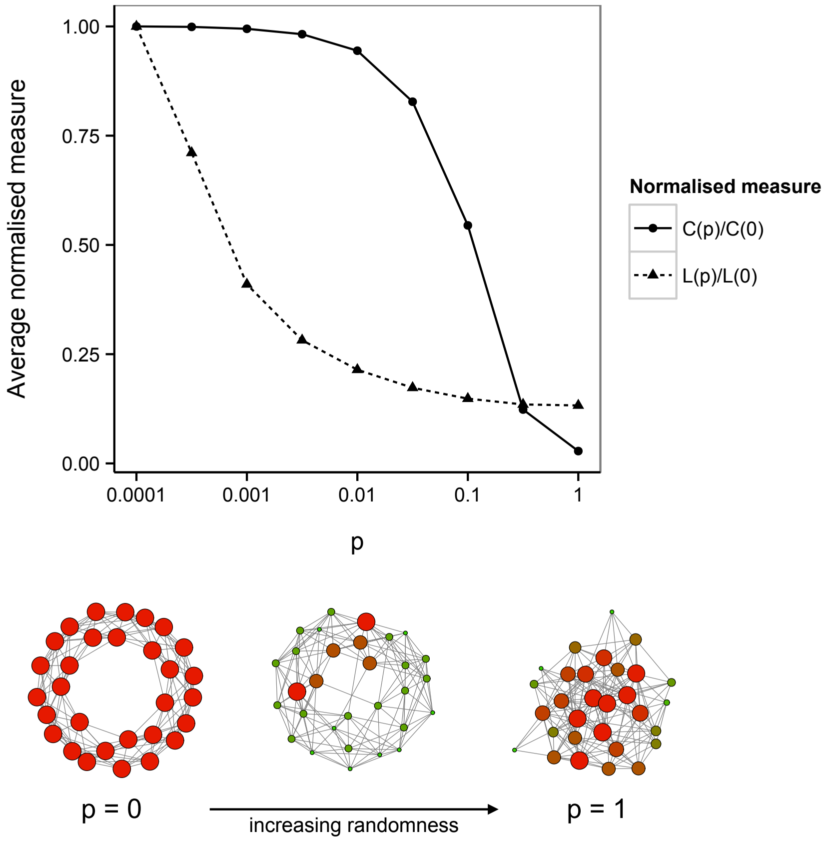


Figure 6.1 – The relation between average path length (L) and the clustering coefficient (C) of a network is shown in the line graph. Each data point represents the average of 20 values for L and C, obtained by randomly re-wiring the connections in the networks with probability p. Values were normalised on $p = 0$, the number of nodes was $n = 30$ and the neighbourhood was $k = 5$. See text for details.

In order to test whether the scientific record of terms used in association with ætiologies of developmental dyslexia represent a network that is organised as an optimised knowledge base, a small-world test was conducted (cf., Watts & Strogatz, 1998). The same corpus of 1407 abstracts reported in Chapter 1 was used, now including the abstracts dating to 2013 (April), after removal of common words and words with a frequency of occurrence of 1, words were stemmed and completed using various text mining packages in the R software environment (see the online supplementary information for details). Table 6.1 shows the resulting number of abstracts for each epoch, the sparsity of the term-document matrix (number of zero elements on total elements in the term-document matrix) as well as the number of terms (n, nodes or vertices), and the average degree of each node (k, the mean number of edges of a node) in the networks. The networks are shown in Figure **6.2** and Figure **6.3**.

The test for small-world structure concerns a comparison of the average path length (L) and the average clustering coefficient (C) of the actual network to a randomly rewired version of the network with same number of nodes and average node degree. The probability of rewiring two connected nodes was set to 1 and the values were normalised by dividing the values of L and C for re-wiring with probability 0 (see Watts & Strogatz, 1998 for details). Table 6.1 shows that the networks of the decades 1970-1980 and 1980-1990 can be classified as small-world networks, the other networks truly represent a weak knowledge base: Given the number of nodes and degree, the actual network is identical to a network in which all the connections are randomly re-wired. The composition of vertices and edges is irrelevant and could just as well have been any other random pattern of connections. Importantly, it is not the number of terms that changes over the years, but the degree with

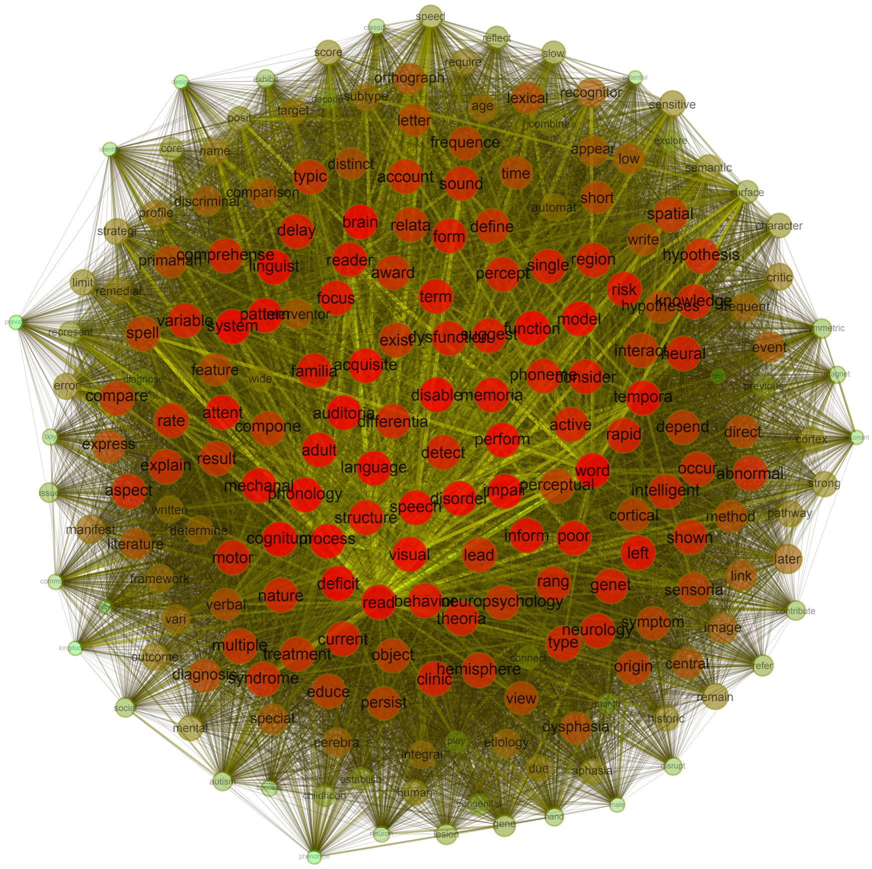


Figure 6.2 – The network of terms ($n = 192$) mentioned in abstracts ($N = 298$) on the ætiology of developmental dyslexia in a period of 2.4 years since 2010. Each term has on average $k = 117$ connections to other terms in the network. The network does not reveal a small-world structure: Compared to a randomly rewired version the clustering coefficient $C(p)$ and average path length $L(p)$ are the same as the original. See Table 6.1 and text for details

which nodes are connected to each other (around 200 nodes in all the networks, these numbers fluctuate because of the trade-off in trying to keep the sparsity level approximately equal). Most striking perhaps is the network of Figure **6.2**, representing just 2.4 years of the recent literature. It is already almost fully connected.

6.3 Realism revisited

Some of the suggestions put forward, like exquisite context sensitivity, may be discomforting to an empirical scientist. If not measurement outcomes, what should we be realist about in soft empirical science?[3] A scientific realist believes that our best scientific theories do not merely save the phe-

[3]This section contains excerpts from an unpublished manuscript: Hasselman, Seevinck, Cox (2011).

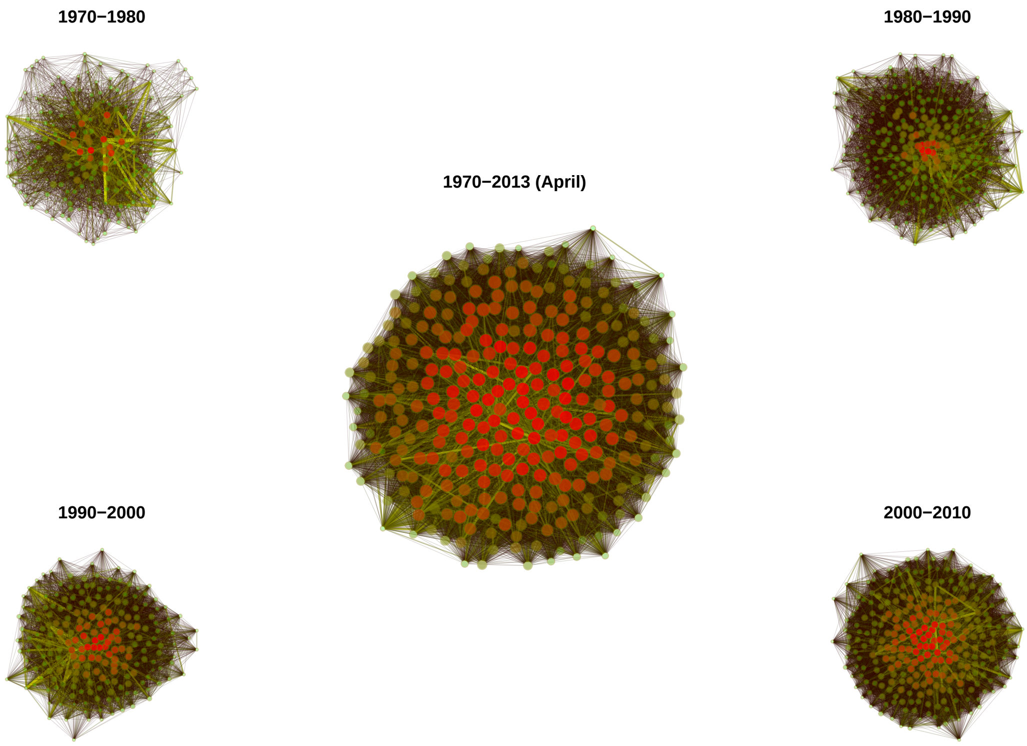


Figure 6.3 – Networks constructed the same way as Figure **6.2**, but for different time periods. The central graph is based on the entire corpus of abstracts ($N = 1407$). See text and Table 6.1 for details.

nomena, but rather that it is reasonable to believe that they capture, in an approximate way, the inner structure of the universe. It may have become clear that it should not be the measurement outcomes or ontology we should be realist about. The history of science teaches us one certainty: Ontology will be proven wrong, awkward, silly, or downright preposterous eventually (Humours; Phlogiston; Gemmules; the Id, Ego and SuperEgo; Vis Nervosa; Heat and Cold Particles, etc.). There is a way to reconcile the lessons from the history of science and the modesty about what scientific theories tell us about reality. Worrall (1989) presented structural realism (SR) as: "the best of both worlds", as capturing the main pro-realist argument (the no miracles argument) while countering adequately the main anti-realist argument (the pessimistic meta-induction, i.e., the argument from the history of scientific revolutions). The no miracles argument runs as follows:

> "[...] some scientific theories enjoy enormous empirical success; if these theories are not even approximately true, their success is miraculous; on the other hand, if these theories are approximately true, their success is not miraculous; we should thus infer that such theories are approximately true." (Ainsworth, 2010)

At the basis of the pessimistic meta-induction argument is the fact that there has been ontological discontinuity across theory-change (French & Ladyman, 2003). The argument lists occurrences of such changes in the history of science and concludes by induction that it is most likely that the ontology of our currently accepted theories will also be radically revised sooner or later. This is taken to undermine scientific realism. For example, Fresnel's wave theory of light was, by any standards, empirically successful: It correctly predicted various surprising results, for example, the existence of a white spot at the centre of the shadow that is cast by an opaque disc that is held in light coming from a point source. This theory presupposed that light is a wave transmitted by an all-pervading mechanical medium, called the luminiferous Æther discussed earlier. Since physicists no longer believe that such a medium exists the theory appears not to be even approximately true (Ainsworth, 2010).

Worrall (1989) proposed SR (in its epistemic form) as a means of reconciling these two arguments (early versions of structural realism can be found in the works of Russell, Poincaré, and Carnap). The idea is that 1) we should believe what scientific theories tell us about the structure of the unobservable world, but 2) we should be sceptical about what they tell us about the posited ontology of the unobservable world. Thus, for example, "[…] *it seems right to say that Fresnel completely misidentified the nature of light, but nonetheless it is no miracle that his theory enjoyed the empirical predictive success that it did; it is no miracle because Fresnel's theory, as science later saw it, attributed to light the right structure.*" (Worrall, 1989, p. 117). Worrall claims that Fresnel's theory attributes to light the right structure because Maxwell's theory of electromagnetism (which succeeded Fresnel's) retained Fresnel's equations. Fresnel's theory attributed to light the wrong nature because the interpretation of these equations is altered. The alleged mechanical nature of light was indeed replaced by the field-theoretical one (i.e., electromagnetic). In general, Worall claims that with respect to such revisions in ontology, "*[t]here was continuity or accumulation in the shift, but the continuity is one of form or structure, not of content*" (Worrall, 1989, p. 117).

Interpretation is again shown to be irrelevant for theory evaluation. It is important that everything happened according to Fresnel's equations and not what scientists imagined those physical quantities to represent in reality, as direct experience. But this history is not just about theory evaluation… this is unification! There is a crucial role for interpretation here, not yet discussed. When does one raise a conjecture to a postulate? It's not just an intuitive leap; it is an inspired intuitive leap. It happens when scientists start to notice that a particular interpretation of a theory seems to generate new predictions or allows explanation of known anomalies to current theory. When something like that happens, a change of formalism and ontology and indeed unification is imminent! Getting hung up on one version of reality certainly does not increase the chances of such a leap ever occurring. Einstein prescribed the following. A scientist:

> "[…] must appear to the systematic epistemologist as a type of unscrupulous opportunist: he appears as a realist insofar as he seeks to describe a world independent of the acts of perception; as idealist insofar as he looks upon the concepts and theories as free inventions of the human spirit (not derivable from what is empirically given); as positivist insofar as he considers his concepts and theories justified only to the extent to which they furnish a logical representation of relations among sensory experiences. He may even appear as a Platonist or Pythagorean insofar as he considers the viewpoint of logical simplicity as an indispensable tool of research." (Einstein, 1949, p. 684 emphasis added).

To build a theory of principles and raise a conjecture to a postulate, scientists must be unscrupulous philosophical opportunists and stretch their mind and creativity to explore and be inspired by the perspectives of the many different interpretations of reality philosophers offer them. That is why I believe science is one of the arts studying the structure of reality and not an industry that produces facts about reality.

This realism about structure allows for a modest scientific realism. What makes structural realism stand apart from other varieties of realism is that it is more cautious in its realist claims, so as not to fall prey to the pessimistic meta-induction over the history of science, yet at the same time, to remain sufficiently substantive to provide a basis for the no-miracle argument. The only explanation for the fact that in sequences of successively accepted theories each theory is empirically and technologically at least as, and generically more, successful than its predecessor is that they latch onto the structure of the physical world better and better. Modern day versions of structural realism assert that all we can know about the world is its structure (epistemic variety), or that structure is all there is (ontic variety). For the purposes of this book it is not important to adopt one of these two variants of structural realism (there are more!); the important take home message is that we should focus on structure and not get hung up on ontology.

An Inflation Theory of Theory Inflation: Big Bang or Big Crunch?

> "A difficulty of much psychological theorizing is vagueness in the terms employed. In this work, the above ideas have been studied in mathematical form throughout, the definitions and proofs being given corresponding precision." (Ashby, 1945, p. 13)

The weak knowledge base of theories of developmental dyslexia is likely indicative of a more general problem of soft empirical science. It illustrates the scientific cycle has turned into a vicious cycle. It would require more detailed study, but perhaps the increasing saturation or paralysis of the scientific record as evidenced by the network analyses could be used as a predictor of crises of confidence in a field of science. The fact that such crises appear to recur with some frequency could point to a mechanism in which a certain domain in reality specified by a finite set of common language terms such as the ±200 terms in the previous section become fully connected. The number of possible theoretical accounts inflates as the number of connected nodes in the network increases. At some point, there is just no ordered structure left in the network and this causes indifference with respect to testing ontology, discontent with the apparent triviality of theoretical claims is expressed (Dunnette, 1966; Ferguson & Heene, 2012; Ioannidis, 2005; Meehl, 1990; Ring, 1967) and quite possibly opportunities to be sloppy, either by accident, due to pressure, or on purpose, arise (Fanelli, 2009; John, Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011; Simonsohn, 2013). There are three routes out of this predicament:

1. The Big Bang route - Allow theory inflation to continue by domain inflation
2. The Big Crunch route - Stop theory inflation temporarily by rigorous pruning
3. Eternal Chaotic Inflation route - Stop theory inflation temporarily by theory unification.

Route 1 can be observed when new technologies are developed that allow connections to new domains (e.g., availability of neuroimaging techniques or genetic sequencing). Attempts to weed out the uninformative connections of route 2 can be seen in contemporary efforts to drastically reform scientific practice by imposing stricter rules for scientific inference including the imminent replicability Apocalypse (e.g., Asendorpf et al., 2013; Open Science Collaboration, 2012). After the pruning is done, if no formal system of theory evaluation is put in place, there will be a similar situation after a couple of years without having made much advancements in uncovering the structure of reality. Route 3 appears to be reserved for the hard sciences, although unification attempts are sometimes announced, like the unification of neural network models, dynamical models and the complex system approach (e.g., Kloos & Van Orden, 2009), these appear to be local efforts, in this case even a subfield within developmental psychology. Unification does not get rid of theory inflation, which will just continue after a successful attempt, but it does promise to work in the service of uncovering more of the structure of the universe.

Given this predicament, what work can SR do for the advancement of the soft empirical sciences? Some authors would argue that the reductionist approaches of the physical sciences (route

3) cannot be applied to psychological science and that the current theoretical diversity is fundamentally linked to the field, something we should accept as inevitable (McCauley & Bechtel, 2001), just like representationalism and mechanistic theories (Bechtel, 2009). This position is known as explanatory pluralism (Dale, 2008; Dale, Dietrich, & Chemero, 2009) and I interpret it to promote a utilitarian approach to theory evaluation where there can be a peaceful coexistence of theories (ontological indifference) until one theory has proven to be more useful than others. I will not discuss this position here, but kindly disagree with this program and promote the SR approach: Ontological commitment, the kind that Lashley practiced, should be the driving force behind theory evaluation. This can only invoke a fear of reductionist fundamentalism and blinding monism, if it were done without the realization that ontology is only a temporary vehicle for understanding. The explicit goal of ontological commitment is to crash-test theory and ontology against reality. We can only learn about the quality of our explanatory vehicles, the structure in reality they may have latched onto, when they break down. Then we should proceed to build new ontological vehicles, maybe using the parts that survived the crash-test, or when they are fundamentally flawed, design something that is completely new. The scientific cycle should be like a test bed for ontology, not a trench war or a neutral zone.

The SR approach actually allows for a neutral monism (Stubenberg, 2010; Tully, 1999), which is something very different from ontological indifference. Our theories of psychological and social phenomena need to be about compositional relations (Healey, 2010), not about a compositional hierarchy implied by a strict monism. A number of scientists studying psychological phenomena have argued to adopt this aim of science in the past, perhaps now is the time to review their pleas (e.g., Chemero, 2003; J.J. Gibson, 1979; Kugler, Shaw, Vincente, & Kinsella-Shaw, 1990; Michaels & Carello, 1981; Shaw & Turvey, 1999; Thelen, Kelso, & Fogel, 1987; Turvey & Shaw, 1999). The number of revisions and unifications of theories of principles, not the number of individual confirmations and replications of theories of construction, is what will eventually define the success of the empirical soft sciences. Otherwise a Big Bang or Big Crunch event may wipe out the entire field. Physicist are already invading neuroscience with principled accounts of how order arises in neural networks through random fluctuations and emergent complex neural dynamics at the micro scale (e.g., Chialvo, 2010; Orlandi, Soriano, Alvarez-Lacalle, Teller, & Casademunt, 2013; Tagliazucchi, Balenzuela, Fraiman, & Chialvo, 2012). The dynamics at the meso-scale have already been studied in more detail (Freeman, 2000, 2009; Tognoli & Kelso, 2009), so a statistical mechanics connecting the levels is a conceivable avenue to pursue (Lebowitz, 1993; Schweitzer, Ebeling, & Tilch, 2001). In the meantime the “inventors” have pretty much solved the automatized speech recognition problem (e.g., Google translate: Anusuya & Katti, 2009; Bikel & Zitouni, 2012; Chelba, Bikel, & Shugrina, 2012) and the inventors are creating their own campuses and labs to study human behaviour from decision making, behaviour in social networks, to artificial agent interactions (e.g., Spector, Norvig, & Petrov, 2012). What will be left to claim for social science when the dust settles?

6.4 Principles revisited

> ``An idea is not explained if we know that at the time of its occurrence a certain change occurred in a certain part of the brain. […] It may be that at some future time an anatomist can so accurately examine the brain with a microscope that he will be able to say with surety, `This person had such and such sensations, such and such memories,' etc., but he can attach meaning to these statements only by calling up the phenomena to which they correspond in his own mind. […] The science of the changes of molecules which corresponds to ideas is no more the science of ideas than the science of printed words is philology.''
>
> -Scripture(1891, pp. 308-309)

Taking seriously the SR position on ontology, something might be gained if a commonly used observer's perspective were identified and explicated and formally evaluated for its role in incapacitating theory evaluation in empirical social science. One way to find out about such a perspective is to ask what formalism is most commonly used in empirical social science. Perhaps the least controversial answer is that it is a set of theorems and axioms that describe ergodic systems (Molenaar, 2008; Molenaar & Campbell, 2009) and probabilities of random events. Assuming the object of study is an ergodic system is rather convenient as it opens up the possibility to use the laws of probability to the predict properties of the system: Ensemble statistics (central tendency measures) represented as the expected values of random variables. Random processes describe change in ergodic systems in which future states are not, or only slightly influenced by the history of system states. Most models of change for complex living systems concern processes that are essentially deterministic, where future states are completely defined by the history of the system. The dynamic system theorists and the proponents of the complex systems approach indeed argue that the ergodic condition (Kalikow, 2011; Petersen, 1996) does not apply to living systems (Holden, Choi, Amazeen, & Van Orden, 2011; Van Orden, Kello, & Holden, 2010).

The classical distinction of the statistician between randomness and order is the dimension of indeterminism (Neyman, 1960). Assuming apparent indeterminate fluctuations are independent and random, means studying the change of ergodic systems over time is essentially not very different from observing many systems of different 'ages' at one point in time. The same logic holds for the repeated measurements of one observable of an ergodic system: Throwing one die 100 subsequent occasions gives the same expected measurement outcomes as throwing 100 dice on one occasion. That is why social science, most of the time, collects data in samples of human beings and calculates properties of the sample that are generalized to reflect properties of the population. The first sentence in the quote by Meehl in section 6.1.2 reflects such an ergodic assumption for disease: "[...] *a cluster of symptoms, complaints, and signs that covary over the population of patients (and usually over time in the individual patient)*" (Meehl 1977, p. 51).

The term "usually" hints that things may be more complicated in reality. The 'technology' of Social science applied to reality is about groups in a society, individuals in groups, a patient in a dyad, a consumer in an economy, a student in a curriculum, an employee in an organisation, a developing child in a family, or an aging grandmother in a home for the elderly. To be able to attribute properties of the ensemble (the population), to properties of the individual (= statistical syllogism), the ergodic assumption is essential, but almost certainly false when applied to living systems that are open to interact with the dynamical environments just described (e.g., Kievit, Frankenhuis, Waldorp, & Borsboom, 2013). The mathematics of random variables suggest that even when the ergodic conditions are met, property attribution to individuals may not be possible in principle (Ellis & Van den Wollenberg, 1993).

6.4.1 What's Higgs got to do with it?

It is likely that the component-dominant ontology of mechanical systems with Newtonian dynamics, together with erroneous use of statistical syllogism based on the assumption of ergodicity has caused the differences between ensemble properties of populations to be wrongly attributed as actual deficits in individual dyslexic readers. The meta-theoretical causes of this error were discussed in Chapter 1 and they are associated to the causes mentioned in this chapter. The statistical syllogism is logically invalid if the ergodic condition does not apply and the presence of anything remotely similar to a crud factor would invalidate the ergodic condition.

Nonetheless, the ontology of complex systems in which interactions dominate in dynamically "causing" the behavioural state of the whole was suggested to be able to accommodate an explanation of observed component-like behavioural modes in section 6.2.1. As shown in **6.4**, there seems to be a distinction between observed behavioural modes (or rather, their interpretation) along the "dimensions" of component versus interaction dominance and indeterminate (random)

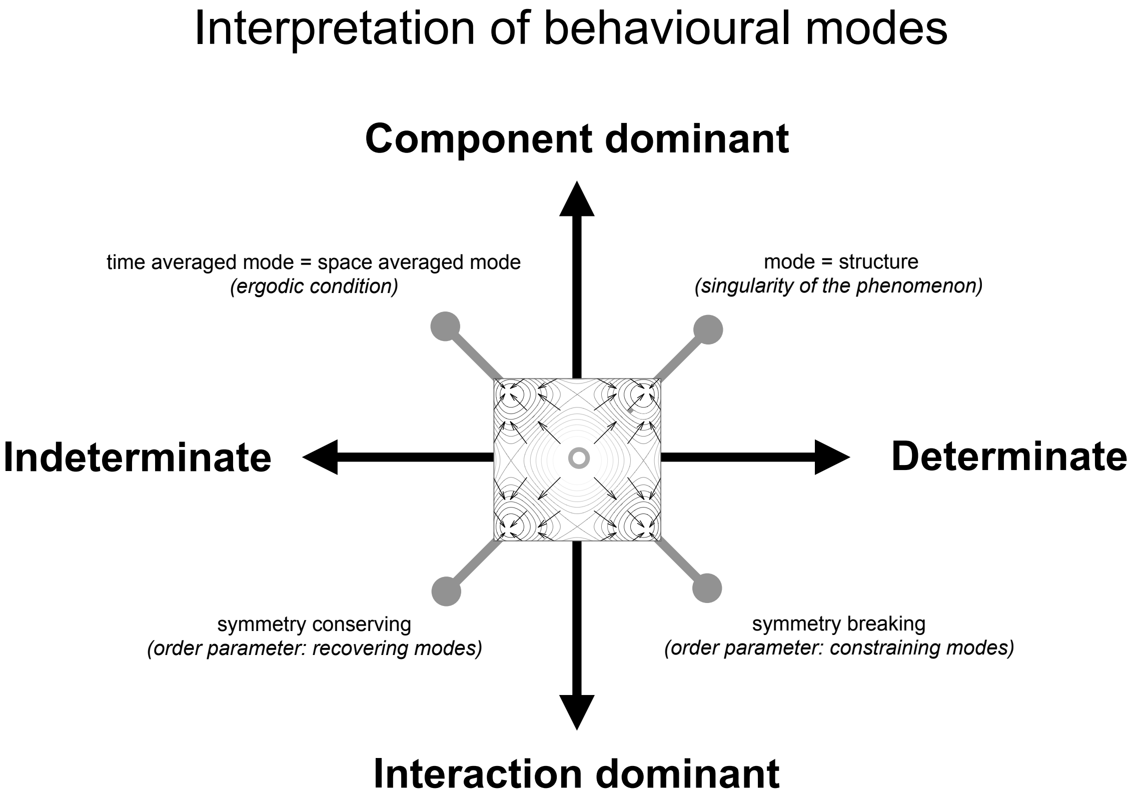


Figure 6.4 – Interpretation of behavioural modes. The component-dominant part (top) of the figure may be due to static observation of behavioural modes of the same system whose modes are observed dynamically in the interaction-dominant part (bottom) of the figure.

versus determinate behaviour (ordered) modes of behaviour. In the upper part of the figure, the uniformity of the behavioural mode is the prime object of study. If a system is interpreted to behave determinate component-like, sources of individual variation tend to be uniform, or attributable to a source or a factor. The module mistake is often made in neuropsychological case studies and a component architecture for cognition is inferred (Van Orden & Kloos, 2003; Van Orden, Pennington, & Stone, 2001). If a behavioural mode is indeterminate component-like, individual variation is substantial, but this nuisance noise can be "cancelled" by averaging, and the remaining variation at the ensemble level can be attributed to a source or a design factor. The ergodic condition is often incorrectly assumed and inferential statics are inappropriately used for induction of properties of individuals in the ensemble, who are basically thought to be identical like an ensemble of particles.

In the interaction-dominant part of the figure the change of the behavioural modes over time are the objects of study. This implies that the theories based on the component-dominant ontologies could be looking at snapshots of singular behavioural modes of the very same system the interaction-dominant ontology studies (as suggested in paragraph 6.2.1). Interaction-dominant determinate behaviour is symmetry breaking in the sense that many possible behavioural modes are reduced to just a few. This symmetry breaking behaviour is what the potential models discussed in Chapter 5 describe. In fact, spontaneous symmetry breaking behaviour is a fundamental property of physical theories; the "Higgs mechanism"[4] describes spontaneous symmetry breaking of a so-called "Mexican hat potential" that emerges in a high-energy context. The potential state in the middle (top) of the Mexican is unstable and a cross section looks exactly like the multi-stable po-

[4] An excellent explanation of this mechanism, including helpful illustrations can be found here: http://www.quantumdiaries.org/2011/11/21/why-do-we-expect-a-higgs-boson-part-i-electroweak-symmetry-breaking/

Box 6.1: A formalism is not a theory, but a set of postulates and axioms (not necessarily stated in terms of a formal calculus).

A formalism defines:

1. An explanatory domain in reality
2. The relevant phenomena for scientific explanation
3. The systems in which the relevant phenomena may be observed
4. The scales at which degrees of freedom can be identified that are involved in generating phenomena
5. Any rules for assigning properties to posited entities (measurement, or interpretation of measurement)

The formalism marks a specific slice of reality as the arena for scientific theories to compete for precision and accuracy of their predictions about the phenomena of interest.

tential in Figure **5.4** (middle row, $k = 0$). In most explanations of this physical phenomenon a ball metaphor is used, just like in Chapter 5. The ball rolls from the centre of the multi-stable potential field into the deeper potential well in the rim of the hat, due to random fluctuations. In principle this describes the same kind of behaviour as the models discussed in Chapter 5, except of course for the much more detailed connection of the parameters of the potential field to theoretical entities of particle physics corroborated by highly accurate measurements.

The dynamical behaviour in which symmetry is recovered and more potential behavioural modes become "available" is less known, or perhaps not used as often as a metaphor for complex adaptive behaviour, but it is of course equally important in order to escape determinacy and get stuck in a single behavioural mode. It is essential to understand development, intervention, learning and plasticity. Many of the mathematical dynamical systems models that can display deterministic chaos have been described to reveal symmetry conserving behaviour for certain parameter settings, in other words a range of potential behavioural modes lost due to symmetry breaking are recovered for those parameter settings (e.g., Chossat & Golubitsky, 1988; Kugler & Shaw, 1990; Lim & Kim, 2001). Simulations of complex networks show that only very specific topologies of organization reveal a property called extended criticality; a hierarchical organization of clusters of interconnected scale-free networks (Kaiser, Görner, & Hilgetag, 2007). Moreover, results show that increasing the complexity of the network structure is associated with increase in optimality of performance (Kaiser & Hilgetag, 2010).

6.5 Driven By Improbability: A Formalism For A Physical Science of Human Nature

> "The Infinite Improbability Drive was invented following research into finite improbability which was often used to break the ice at parties by making all the molecules in the hostess' undergarments leap one foot simultaneously to the left in accordance with the theory of indeterminacy. Many respectable physicists said they weren't going to stand for that sort of thing, partly because it was a debasement of science, but mostly because they didn't get invited to those sort of parties."

-Douglas Adams, Hitchhikers Guide to the Galaxy

The lesson to be learned from the physical interpretation of potential fields, phase transitions, and the dynamics of complex systems is that it is possible to go beyond the metaphor when applying these mathematical tools to theorise about human behaviour in the empirical social sciences

Box 6.2: Terminology commonly used in empirical social science to describe complex human behaviour, defined in terms of the behaviour and observable properties of open, complex dynamical systems common in (statistical) thermodynamics and systems biology. This reconception can provide the basis for a unifying formalism for order generation in biological systems, grounded in physical law.

Behaviour: An observable response to changes in the external environment or constituent subsystems of a complex system.

Information: A measurable quantity that resolves the improbability or uncertainty of an event or system state. System states that are highly uncertain, improbable or rare, require more information to describe than probable states. Information can only distinguish one thing from another within the context of uncertainty and probability definitions. Information theory cannot help distinguish meaningful from meaningless (biological) information.

Entropy: Expressed as a quantity of information, it is the amount that is needed to fully describe the system's micro-states, or the degrees of freedom it has available for behaviour. The entropy of a thermodynamical system is maximal when there is a complete absence of order in the distribution of its energy, that is, its internal structure will be in a state of maximal disorder and require maximal information to describe.

Order Generating Process (Computation): Any process or system configuration that increases the order (= decreases the entropy) of the internal structure of a system. An OGP changes the amount of information needed to describe the states of the system. Applied to behaviour of biological systems, prediction, anticipation, self-organisation, information-processing and adaptive behaviour all indicate an OGP, or order transitions between stable modes of behaviour (Order Parameter Dynamics).

Representation: Either a physical record of the states of a system (e.g., state space trajectories visualised after a measurement procedure; growth rings in a cross-section of a tree trunk), or, the reproduction of system states, as a repeated presentation to an observer. A harmonic oscillator re-presents its amplitude peak at a specific frequency, mammals re-present stable behavioural modes synchronised to changes in their external environment caused by the celestial mechanics of sun, earth and moon.

'Adaptive' behaviour: Any behaviour that can be conceived of as resulting from the 'after-effects' of an interaction event, an order generating process.

in general, and to ætiologies of developmental dyslexia in specific. Based on principles of composition and dynamic interaction common to physical, chemical, and biological systems the empirical social sciences have made an important theoretical advancement that is often overlooked due to epistemic sloughing by hypothesis testing. The fact that predictions based on interaction-dominant ontology concern shapes of functional relations and point-range predictions instead of mere signs of associations, that is, specifications that venture into Stage C of Box **1.4**, is an important step towards proposing theories of principles and testing theories in the context of strong inference (Hasselman, 2013). Chapter 4 provides strong arguments for a Complexity Matching Hypothesis of perception and action that is based on the multifractal spectrum of speech sounds. Many of the studies departing from the (multi-)fractal formalism of behaviour (Ihlen & Vereijken, 2013; Stephen, Anastas, & Dixon, 2012) and the theoretical framework of fractal physiology of health and wellbeing (Eng et al., 2002; Van Orden et al., 2009) predict a range of values for scaling parameters (or fractal dimension). For instance, predictions about proficient and healthy performance based on power spectral density reveal power-laws close to -1 (1/f noise). This is a very risky prediction; the measurement outcomes cannot be 0 or -2 as this would indicate dynamics that are associated with impaired behaviour or disease. evidence for a role of framework has been shown to apply to developmental dyslexia (Wijnants, Hasselman, et al., 2012; Holden et al., 2015) and the corroboration strength of these point-range predictions should add considerably to the verisimilitude of the interaction-dominant ontology.

In a similar fashion the physically inspired prediction of a specific functional relation, or "fingerprint" of a phase transition that should be observable in the entropy of measurement outcomes is a risky prediction that has been shown to be empirically accurate and precise (e.g., predicting benefits of behavioural intervention, Lichtwarck-Aschoff, Hasselman, Cox, Pepler, & Granic, 2012; emergence of insight in problem solving, Cox & Hasselman, 2009; Stephen, Dixon, & Isenhower, 2009). Even so, there is nothing to prevent other researchers to demand their playground should be used to uncover knowledge about the world, as indicated in the first paragraph, by merely declaring the ontology inappropriate for scientific explanation.

Therefore, a much more profound, and perhaps surprising unification should be attempted between the 'information' theories and the complex systems approach: If the consensus definitions of information, computation and adaptive behaviour are adopted from the physical sciences, a formalism can be proposed (see Box **6.2**) that would place both perspectives in the same ballpark, forcing them to compete to explain the same phenomena. Box **6.2** concerns a redefinition of terminology commonly used in the empirical social sciences in terms biology, information science and physics. It should be noted that these definitions are not derived from consensus formalism for all living systems, because such a thing does not exist. They do express a consensus categorization of complex systems into thermodynamically probable and improbable entities. Living systems appear to (locally) defy the second law of thermodynamics, which states that a closed system will always evolve towards an internal state of maximal entropy (disorder). This attraction to the state of thermodynamic equilibrium provides an entropic arrow of time.

Conversely, any process of biological development or learning and therefore, memory and 'representation' as defined earlier, implies an increase in the complexity of the internal structure of a system. The biological arrow of time (Walker, 1972) runs opposite to the entropic arrow of time and introduces a local violation of the second law of thermodynamics (see Figure **6.2**). Living systems are thermodynamically highly improbable entities (Schrödinger, 1944), they are often described as dissipative systems (e.g., Lintern & Kugler, 1991; Prigogine & Nicolis, 1977; Prigogine & Stengers, 1984; Schweitzer et al., 2001). Dissipative systems extract matter or energy from their environment in order to increase or maintain the complexity of their internal structure. When order is created in a system using free energy, heat (energy in disorder), is dissipated back into the environment, thus increasing the entropy of the universe.

As shown in Figure **6.6** living systems that defy the universal law of ever increasing entropy are aggregations of broken symmetries, a nested causal structure (Hopfield, 1994). The change in internal structure implies that the system becomes increasingly specified with respect to events in the environment (compare the ontological descent described by Shaw & Turvey, 1999; Turvey & Shaw, 1999). Living systems of the same kind loose their identity, whereas lifeless particles of the same kind share identity. If the internal structure of one particle is known, all of them are known, including their expected path through a field. To know a living system and predict its future behaviour, one has to learn about its specification to its environment, but also to past events that changed the internal structure of the system and made it an increasingly improbable system from the perspective of thermodynamics.

6.5.1 Order Generation And Information

Equating intelligent behavior to Computation as logical structure as is common in any science that uses computational models to simulate intelligent behavior can only be an *interpretation* of its function, to claim a computational architecture, virtual machinery for the mind, capable of simulating some cognitive phenomenon must also be physically realized in the brain would be committing the *interpretation fallacy* of chapter 1. Boyle lucidly explains the unrealistic goals of the 'interpretative view' ofbiological computation:

> ``This interpretive view of computation is responsible for the widespread use of functional models to understand cognition and computer programs to simulate mental behavior. Such

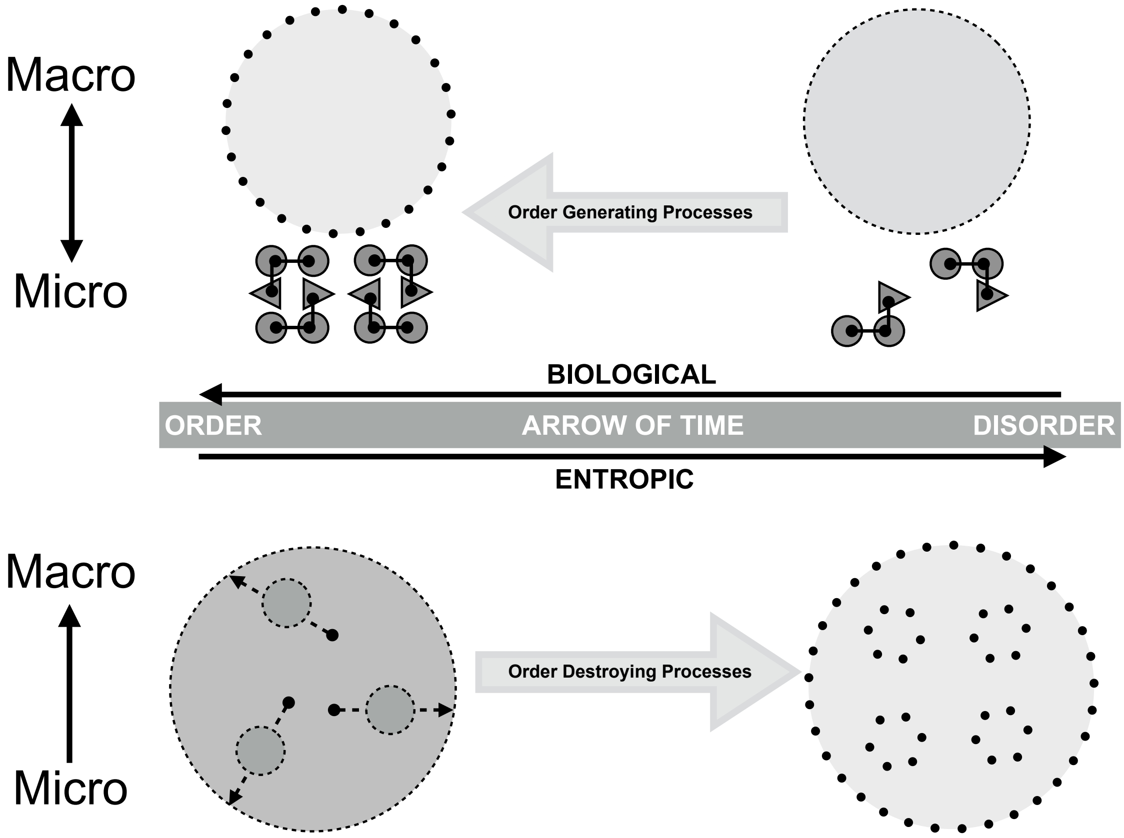


Figure 6.5 – The Biological and the Entropic arrow of time. In Biological open complex systems the internal structure is increasingly specified due to Order Generating Processes. In general, order transitions imply a change in entropy and information, relative to the previous state, a higher order state will be less entropic, meaning less information is needed to describe its states. In non-biological open systems, when no energy is added to the system, each change in structure, e.g., due to diffusion, will reflect a loss of internal structure. In general Entropy will increase, therefore the information needed to describe the states of the system will increase.

> models, however, fail to tell us anything about physical characteristics of the brain's information processing, only that we can interpret the brain as computing some class of functions. Yet it is from those physical characteristics that minds emerge, underscoring the importance of understanding the basis of this phenomenon. Certainly we won't understand it by identifying the class of functions the brain can be interpreted as computing, nor is that even a remotely realistic goal given the complexity of the brain and our behaviors." (Boyle, 1994).

From a biophysical perspective, the most important false contradiction between the information-based or constructive theories of human behaviour and the complex systems approach involves the use of the words computation and information applied to adaptive behaviour (see Box **6.2**). The ability of living systems to defy a genuine universal law of physics, even if the insurgency is only local and temporary, is an extraordinary phenomenon. At the heart of the matter is the realisation that there are change processes and system configurations that increase the complexity or order of the internal structure of a complex system. How such processes are modelled, formalized, or conceived of in a scientific explanation, for example as the result of computation due to a self-organised critical process (cf., Langton, 1990), information- field computation (cf., MacLennan, 1999) or the application of the sequential rules of an algorithm on abstract symbols, is irrelevant. What is important to acknowledge is that all theoretical descriptions of cognition and adaptive behaviour concern the same observable phenomenon: The increase of order in a system, the establishment or transition to a new order in a system, due to an interaction of system components with the order gener-

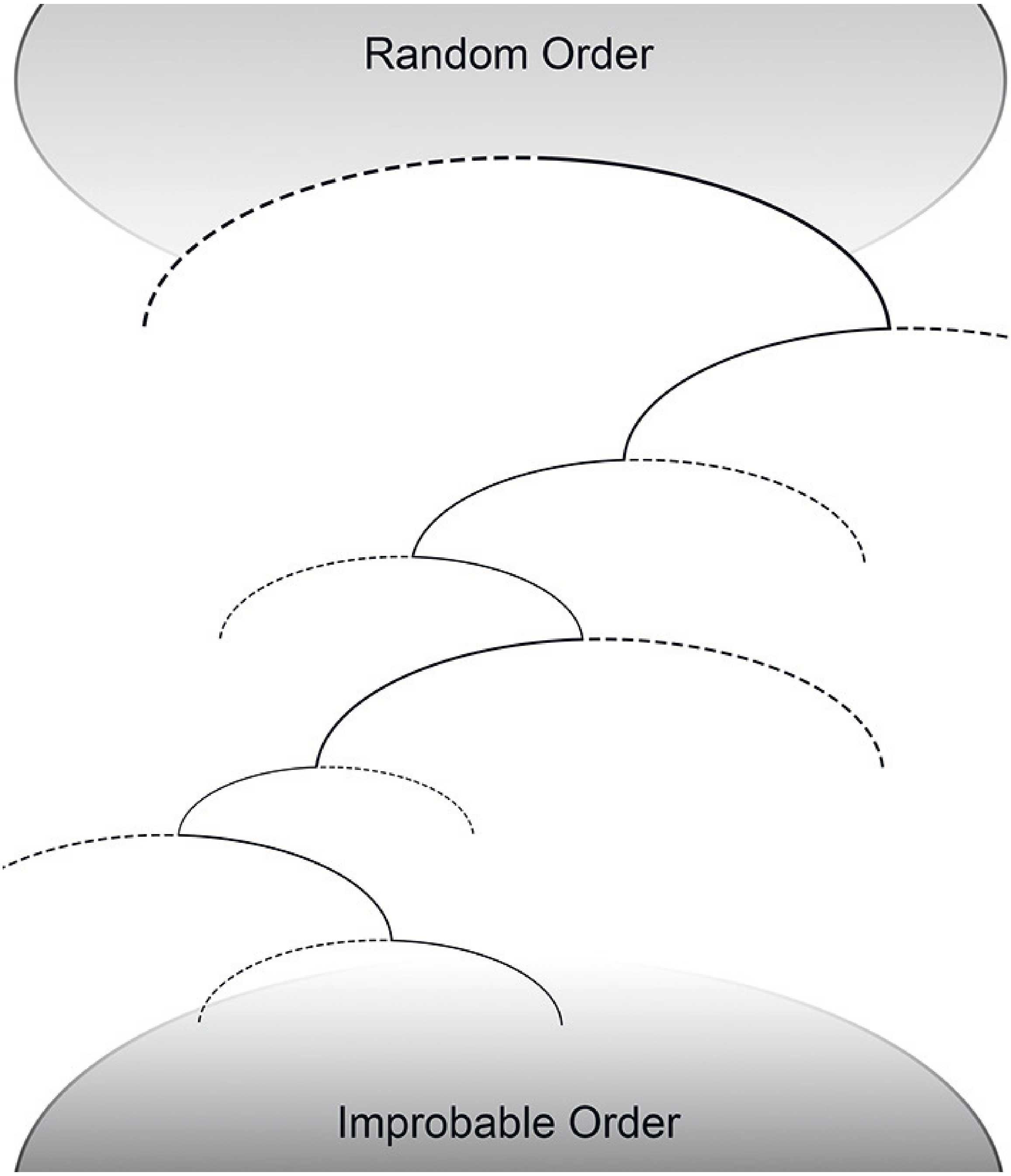


Figure 6.6 – Nested causation as an accumulation of broken symmetries, driven by thermodynamic improbability, or an increase in specificity of the system (compare to the ontological descent from possibility to actuality discussed in Shaw & Turvey, 1999; Turvey & Shaw, 1999).

ating process. Any process that changes the information needed to describe the states of a system against the law of increasing entropy is an act of order generation. It doesn't matter whether you call it computation or self-organisation: The thermodynamic improbability of the system increases as a consequence of the specification of its internal structure to a change in the environment.

This is one of the primary functions of formalism: Generalization of the description of the phenomena of interest, such that theories that may be based on different ontology compete to explain the same phenomena (See Box **6.1**. If it is not possible to produce testable, diverging predictions based on theoretical accounts of such diverse ontology their dissimilarity must be considered trivial for all intents and purposes, nothing can be considered 'vague' or 'exotic' if all constructs can be shown to claim explanatory power over the same phenomena. This is what the quantum formalism achieved for the description of light and matter in terms of either particles or waves: The descriptions were found to be equivalent from the perspective of the quantum formalism (Heisenberg, 1971).

A first step to achieve a unifying formalism for psychological science must be to introduce the term order generating process (OGP) and demand it be defined and specified by competing theories. Whether the claim is that biological systems compute, construct, or self-organise order, a local violation of a universal law of physics is implied that needs to be accounted for. One cannot just conjure order into existence to explain a systematic pattern in the empirical record. The use of self-organization (with or without criticality) as an OGP has a strong plausibility advantage over computation as a logical structure, as it is known as an uncontroversial change process in physics, biology, robotics and computer science (Aschwanden, 2011).

6.5.2 Re-Presentation = Re-Production

> "Our understanding of biological computation and its origins must come through studying the relation between computation and its underlying hardware, not computation as a logical structure."
>
> —Hopfield (1994, p. 56)

When a comparison is made between representation in physical systems and all the varieties of mental representation that are frequented in the literature, physical representation in fact simply refers to adaptive behavior that can be *re*-presented by the system. This re-presentation concerns the observed recurrence or reproduction of states or modes of behavior, structural configurations, or complex temporal patterns observed in system observables. Re-presentation of behavior in physical systems does not require any trace of the temporal dynamics of the original behavior to be stored in the internal structure of the system. For example, oscillators reproduce states without retracing an internal record of change, celestial mechanics reproduce circadian and seasonal states of the environment due to the laws of physics, to which most living systems adapt their behaviour.

A useful notion to understand how adaptive behavior of complex systems to events fluctuating on different timescales is achieved is the adiabatic separation described in Box **6.2** (Hopfield, 1994). When the environment of a complex system changes over time (e.g., it is a dynamic environment) a separation into fast and slow changing system observables can be made. The observables that change on a time scale faster than the fluctuations in the dynamic environment take the slower changing observables as parameters settings. The slow variables will fluctuate at the pace of changes in the environment and upon such a change, the faster processes will adapt to a new parameter setting that is global from their perspective and more or less static. This may prompt a shift to a new behavioural regime or could go unnoticed as a sudden change of order, but register as slow adaptation: Learning, growth, and development. This system property allows for adaptation of the internal structure of the system with respect to time scales indicated by the recurrence of values of system observables, in other words, the adaptation is relative to a certain scale of fluctuation. The motion of a complex lifeless system that appears to adapt its internal structure due to adiabatic separation is described as 'adaptive motion' (Hopfield, 1994), to contrast it with the motion of a classical particle in a field that is completely determined by the laws of mechanics and the identity of the particle. To know its classical path through a field, we do not need to know about its internal structure. Adaptive behavior is the behavior of a system that appears to be coordinated by events in its interaction history with the environment.

Learning a new task, such as reading, is an adaptive behaviour that takes place relative to a scale of fluctuation of processes and changes in the environment that appear as slow changing parameter settings (collective variables, synergies) to the immediate and direct sensory experiences occurring in the moment of trying to perform the new task. Adaptive behaviour concerns the exploration of a parameter space for the most optimal configuration. The scale of fluctuation relative to which species change identity occurs at such a slow rate, it cannot be observed directly unless a record of changes is available (e.g., a fossil record). Nevertheless, both learning and evolution by natural selection are order-generating processes that can be described as the adaptive behaviour of a complex system leading to specification of the system to changes in its environment that are reflected by changes in the complexity of its internal structure.

As mentioned earlier, the concept of the mental representation as an information structure (a structure whose description in terms of the amount of information needed to describe its states, is subject to change) appears to be conflated with meaning. Information structures can encode for meaning when they are part of a larger system of interacting components, but meaning does not emerge out of information; weight does not emerge out of kilograms. Meaning emerges out of connections between information structures that are different in nature, or can be said to occupy different 'worlds', by means of a common code (Barbieri, 2006). For example, the behaviour of a very common complex system of organic chemistry known as 'read a genetic code', refers to the phenomenon in which a transfer RNA molecule connects the world of nucleotides (DNA) to the world of amino acids, by means of two separate operations that are called 'recognition' processes (Barbieri, 2003, p. 98). As such, meaning has three ingredients:

1. Information structures of dissimilar identity according to some criterion
2. A code, or codon / analogon pair (Walker, 1983), to connect the structures in 1) across the identity divide
3. Processes of recognition that embody the 'generation' of meaning.

A better way to describe 3) is that codes give meaning to information structures by reproducing similarity by analogy. Walker notes that in the context of reproduction, the following applies to 'analogy':

> "Two separately identifiable patterns are related by analogy if the existence and frequency of the one is correlated with the existence and frequency of the other in the absence of direct forces between the two patterns that could cause the correlation. That is, correspondence between codon and analogon came about, and is maintained, by reproduction of an initial random event." (Walker, 1983, p. 809)

The point here is that if a force would be responsible for the correlation to be captured by the analogy, there would be no need to map the correspondence. The same holds for the arrangement of code components, if some force dictates their order, they cannot efficiently code for anything, except perhaps characteristics of the force itself. This defines organic memory or representation as a very special natural phenomenon capable of accumulating accidental event occurrences (Walker, 1972; 1983). If the base-pair, tRNA, amino acid system is applied by analogy to the neural system, then the brain would either be the base-pair or the tRNA (the analogon) facilitating recognition of systematic correlations between two scales or 'worlds': Perception and action. The system responsible for re-presenting must be larger than the CNS, most likely the body cannot be excluded from the whole, as argued in the chapters of this book. The observation of the re-presentation of a phoneme by a child, for example, when asked to reproduce a letter-sound, concerns the recurrence of a highly specified state of the system that can be modelled as a configuration of parameter settings of collective variables or synergies.

OGPs change the amount of information needed to describe the system, with respect to its previous state. This is not equivalent to meaning. Recognition of systematic patterns in the change of information across lags of time by means of reproduction by analogon could certainly encode for meaning, but the information itself is still meaningless! Both classical and quantum information theory, explicitly exclude meaning as a part of their explanatory domain (Desurvire, 2009). As an example, the total number of possible base-pair sequences that could have constituted the genome of E. coli is $\approx 10^{2400000}$, an unimaginable large amount of information. Yet, much less than 1% of the possible combinations, encode for anything biologically meaningful, that is, "*– correspond to an organism, and enable it to maintain its cellular structure by transforming energy from one form to another, grow, adapt and reproduce*" (Haynie, 2008, p. 347). Information does not specify what it codes for and it cannot be used to distinguish between things that have meaning or not. Again, we must conclude that the only meaningful reduction in the science of behaviour is to relations between things.

Expressed in somewhat archaic but effective language, Walker describes what biological memory function defined in terms of thermodynamic improbability or order generation implies:

> "Memory creates complex structures by mapping complex accidentials, and it accumulates repeats of these complex structures by irreversible copyreproduction." (Walker, 1972, p. 227).

The conception of the brain as a storehouse for frozen information about experienced events begs a plausible physical explanation (Gibson, 1966). Walker suggests that the fundamental property of development and learning is not storage, but the prospective redundant reproduction of thermodynamically improbable states, order generation. As mentioned in Chapter 4, the brain might be a self-tuning, self-affine resonator, in which the multifractal spectrum can be regarded as an anologon, the temporal singularities, the scaling exponent **are** the re-presentation of the temporal structure of the whole.

6.6 Conclusion?

To summarise, the theoretical and empirical arguments presented in the preceding chapters should encourage a program of empirical inquiry based on interaction dominance to study complex disease entities such as developmental dyslexia. Theories and formalism should, as a first sketch, be based on the known principles and universal laws that govern the behaviour of complex dynamical systems. The focus on efficient causes in the form of defective components or component processes such as phoneme representations and phoneme-based auditory processing has been shown to be an unviable direction to advance scientific understanding of such complex ætiology. Both the 'impaired 'and 'normal' phoneme representation concern a description of a highly improbable internal order that accumulated as a history of symmetry breaking events. If each aggregate of nested efficient causes is allowed its own scientific theory, this can only lead to the kind of radical pluralism recently suggested as a multitude of cognitive subtypes of developmental dyslexia related to individual differences in brain structure (Jednoróg, Gawron, Marchewka, Heim, & Grabowska, 2013). What should be studied is the entire history of specification of the internal structure and this comes down to studying adaptive behaviour to symmetry breaking events and symmetry recovering events (or extended criticality) at different temporal and spatial scales of analysis.

As such, this entire book is much more a beginning than a conclusion. Principled theories of the kind produced by the natural sciences are within arms reach of the social sciences and the prospect of being able to shed the qualification of "soft" empirical science must be a motivation to further develop research programs towards appraising and amending theories of principles. There is nothing about human nature that would prohibit a natural science to become temporary realists about its unobservable structure.

A crude sketch of the nomological net [Chapter 1]

How to download and edit the abstracts from pubmed

Table A.1

Table A.2

The following search string was used to obtain records with a downloadable abstract from http://www.ncbi.nlm.nih.gov/pubmed :

> Etiology/Broad[filter] AND ("developmental dyslexia"[All Fields] OR "dyslexia"[All Fields] OR "reading impairment"[All Fields] OR "reading disability"[All Fields] OR "dysphasia"[All Fields] OR "alexia"[All Fields] OR "word blindness"[All Fields] OR "word-blindness"[All Fields] OR "developmental aphasia"[All Fields]) AND hasabstract[text]

After additional filtering (e.g., studies of acquired dyslexia, empty or duplicate records) in a corpus of 1407 documents was further parsed in R+ using the textmining package "tm" (available at http://tm.r-forge.r-project.org). The text was first cleaned from punctuation characters, numbers, English stopwords and very common scientific jargon (e.g., significant, outperformed, control group, etc.). After stemming the remaining words, the stems were categorised into terms that are significant for the study of developmental dyslexia. The first column of Table A.1 lists the terms, if a context was available a term could be extended to be more specific, such as cns~pat to signify a group of words that expressed a pathology of the CNS or per~aud to indicate auditory perception (distinguishable from speech perception ~spp~).

Words that could not be assigned to a term category were deleted. A term-document matrix was created for each decade, listing all the terms in the corpus in rows and documents (abstracts) in columns; the cells are frequencies of occurrence. To create the graphs the term-document matrix was transformed into a term-term matrix, or adjacency matrix. The cells now indicate how often two terms are used in conjunction in the documents of the corpus, this frequency count is the weight that is assigned to the edges of the two connected vertices (terms) in the graph representation of the matrix.

The R+ package"igraph" (http://igraph.sourceforge.net) was used to create the graphs in Figures 1.2 and 1.4 and calculated the degree distribution, co-citation coupling, and graph strength. Annotated R-scripts and raw data necessary to recreate the graphs are available at the open science framework project page for Chapter 1: https://osf.io/8y4sq/.

Table A.1
Meaning of Terms Displayed in Figures 1.2 and 1.3, Including the Normalised Degree for Each Decade.

Term	Meaning	1970-1980	1980-1990	1990-2000	2000-2010
~act~	Action	0.35	0.50	0.44	0.69
~att~	Attention	0.10	0.10	0.42	0.71
~bio~	Biology	NA	0.35	0.52	0.67
~cns~	CNS	0.61	0.68	0.65	0.69
~cog~	Cognition	0.19	0.68	0.73	0.83
~com~	Comorbidity	NA	NA	0.17	0.35
~lan~	Language	0.68	0.75	0.75	0.90
~lit~	Literacy	0.65	0.85	0.83	0.90
~mem~	Memory	NA	0.48	0.52	0.75
~scr~	Script	NA	0.40	0.50	0.63
~spp~	Speech perception	NA	0.35	0.44	0.44
~spr~	Speech production	0.58	0.80	0.79	0.77
~theo~	Theory	NA	NA	0.17	0.62
~trt~	Treatment	0.61	0.78	0.75	0.81
act~spr	Action speech	0.16	0.18	0.10	0.35
bio~bod	Biology body	0.16	0.70	0.60	0.85
bio~dev	Biology development	0.94	0.90	0.96	0.98
bio~env	Biology environment	0.61	0.78	0.88	0.88
bio~epi	Biology epigenetic	0.61	0.50	0.79	0.73
bio~gen	Biology genetic	0.52	0.75	0.75	0.81
cns~ana	CNS anatomical	NA	0.15	0.50	0.44
cns~fun	CNS functional	0.77	0.85	0.90	0.96
cns~msr	CNS measurement	0.29	0.43	0.69	0.83
cns~pat	CNS pathology	0.45	0.70	0.79	0.87
cns~str	CNS structural	0.77	0.98	0.94	0.94
com~aud	Comorbidity auditory	0.19	0.35	0.35	0.46
com~beh	Comorbidity behavioural	0.19	0.58	0.73	0.73
com~cog	Comorbidity cognition	0.74	0.75	0.85	0.67
com~lng	Comorbidity language	0.10	0.28	0.17	NA
com~mem	Comorbidity memory	NA	NA	NA	0.21
com~mot	Comorbidity motor	NA	0.28	0.35	0.38

Table A.1 (Continued)
Meaning of Terms Displayed in Figures 1.2 and 1.3, Including the Normalised Degree for Each Decade.

`com~psy`	Comorbidity psychological	0.61	0.70	0.63	0.75
`com~spp`	Comorbidity speech perception	NA	NA	NA	0.23
`com~spr`	Comorbidity speech production	0.23	NA	0.19	0.21
`com~viz`	Comorbidity visual	NA	0.30	0.23	0.29
`lan~uni`	Language unit	0.23	0.38	0.56	0.71
`lrn~mot`	Learning motor	0.52	0.78	0.71	0.85
`per~aud`	Perception auditory	0.68	0.83	0.77	0.83
`per~mlt`	Perception multimodal	0.26	0.40	0.42	0.58
`per~scr`	Perception script	NA	NA	NA	0.12
`per~som`	Perception somatosensory	NA	NA	0.19	0.33
`per~spp`	Perception speech	NA	0.03	0.25	0.29
`per~viz`	Perception visual	0.65	0.85	0.79	0.94
`scr~tmp`	Script temporal	NA	NA	NA	0.21
`scr~uni`	Script unit	0.42	0.55	0.58	0.69
`spr~lan`	Speech production language	NA	NA	NA	0.12
`spr~tmp`	Speech production temporal	NA	NA	0.27	0.02
`theo~com`	Theory component	0.39	0.55	0.44	0.73
`theo~eti`	Theory ætiology	0.45	0.40	0.63	0.65
`theo~mlt`	Theory multi causal	NA	NA	0.27	0.48
`theo~mod`	Theory model	NA	0.38	0.23	0.38
`theo~mon`	Theory mono causal	NA	NA	0.21	0.52
`theo~sub`	Theory subtype	NA	0.10	0.46	0.65
`theo~tmp`	Theory temporal	NA	NA	0.21	0.60

Table A.2
Meaning of Terms Displayed in Figure 1.4

Term	Meaning
~act~	Action: motor control, balance, speech production, writing, eye movements
~att~	Attention
~aud~	Auditory perception including speech
~beh~	Behaviour: Social development, vocation, self-esteem, personality
~bio~	Biological factors other than CNS and genes: Hormones, physiology, gender, development
~brain~	Structure, function and pathology of the brain
~cog~	Cognition
~com~	Comorbidity
~env~	Environmental factors: socio-economic status, culture, family dynamics
~gen~	(epi-)Genetic factors
~lan~	Language
~lrn~	Learning
~mem~	Memory
~theo~	Theory words: Hypothesis, mechanism, model, theory
~tmp~	Temporal: Rapid naming, slow rise times, fast formants, fluency
~trt~	Treatment: Intervention, ameliorate, therapy, remediation
~viz~	Visual perception
NOTE	Terms removed from the graphs were highly associated with the abstract query: dyslexia, development, reading, spelling, impaired, disorder, deficit. Also some very low frequency categories such as somatosensory perception were removed.

Model Parameters [Chapter 5]

Table B.1

Table B.2

Table B.1

Results for the Fixed Effects of the Linear Mixed Model with Y_{cs} as the Dependent Variable using 15,000 MCMC Resamples. Shown is the Mean across the MCMC Samples (i.e. an Estimate for the Covariate) with Lower and Upper 95% Highest Posterior Density Intervals. The First p-value (pMCMC) is Based on the Posterior Distribution; the Second is Based on a t-distribution with an Upper Bound for the Degrees of Freedom.

Covariate	*MCMC mean*	*HPD95 lower*	*HPD95 upper*	*pMCMC*	*Pr(>|t|)*
(Mean Age, Average Reader, None, Random)	0.192	0.174	0.210	0.0001	0.0001
Age	-0.002	-0.004	-0.001	0.0085	0.0148
Dyslexic Reader	-0.009	-0.033	0.017	0.4820	0.5398
Hysteresis	0.021	0.004	0.037	0.0111	0.0132
Enhanced Contrast	0.057	0.046	0.069	0.0001	0.0001
B>D>B	0.010	-0.006	0.025	0.2084	0.2091
D>B>D	0.056	0.040	0.071	0.0001	0.0001
Age:Dyslexic Reader	0.002	0.000	0.004	0.0252	0.0417
Age:Hysteresis	0.008	0.007	0.010	0.0001	0.0001
Age:Enhanced Contrast	0.003	0.002	0.005	0.0001	0.0001
Dyslexic Reader:Hysteresis	-0.062	-0.084	-0.039	0.0001	0.0001
Dyslexic Reader:Enhanced Contrast	0.007	-0.008	0.023	0.3364	0.3450
Age:B>D>B	0.012	0.010	0.013	0.0001	0.0001
Age:D>B>D	0.010	0.008	0.011	0.0001	0.0001
Dyslexic Reader:B>D>B	0.009	-0.013	0.030	0.4161	0.4092
Dyslexic Reader:D>B>D	-0.065	-0.086	-0.045	0.0001	0.0001
Hysteresis:B>D>B	0.072	0.049	0.094	0.0001	0.0001
Enhanced Contrast:B>D>B	0.040	0.024	0.056	0.0001	0.0001
Hysteresis:D>B>D	-0.044	-0.065	-0.023	0.0001	0.0001
Enhanced Contrast:D>B>D	-0.051	-0.067	-0.034	0.0001	0.0001
Age:Dyslexic Reader:Hysteresis	-0.009	-0.011	-0.007	0.0001	0.0001
Age:Dyslexic Reader:Enhanced Contrast	-0.004	-0.006	-0.003	0.0001	0.0001
Age:Dyslexic Reader:B>D>B	-0.012	-0.014	-0.009	0.0001	0.0001
Age:Dyslexic Reader:D>B>D	-0.011	-0.013	-0.009	0.0001	0.0001
Age:Hysteresis:B>D>B	-0.007	-0.009	-0.005	0.0001	0.0001
Age:Enhanced Contrast:B>D>B	-0.011	-0.013	-0.010	0.0001	0.0001
Age:Hysteresis:D>B>D	-0.005	-0.006	-0.003	0.0001	0.0001
Age:Enhanced Contrast:D>B>D	-0.007	-0.009	-0.006	0.0001	0.0001
Dyslexic Reader:Hysteresis:B>D>B	-0.016	-0.047	0.016	0.3269	0.3253
Dyslexic Reader:Enhanced Contrast:B>D>B	-0.034	-0.057	-0.012	0.0028	0.0031
Dyslexic Reader:Hysteresis:D>B>D	0.209	0.179	0.238	0.0001	0.0001
Dyslexic Reader:Enhanced Contrast:D>B>D	0.074	0.052	0.096	0.0001	0.0001
Age:Dyslexic Reader:Hysteresis:B>D>B	0.012	0.010	0.015	0.0001	0.0001
Age:Dyslexic Reader:Enhanced Contrast:B>D>B	0.013	0.011	0.015	0.0001	0.0001
Age:Dyslexic Reader:Hysteresis:D>B>D	0.010	0.007	0.012	0.0001	0.0001
Age:Dyslexic Reader:Enhanced Contrast:D>B>D	0.011	0.009	0.013	0.0001	0.0001

Note: Number of observations = 44640, Groups: participant ID = 186; Stimulus number = 20

Table B.2

Results for the Fixed Effects of the Linear Mixed Model with n_c as the Dependent Variable Using 15,000 MCMC Resamples. Shown is the Mean across the MCMC Samples (i.e. an Estimate for the Covariate) with Lower and Upper 95% Highest Posterior Density Intervals. The First p-value (pMCMC) is Based on the Posterior Distribution; the Second is Based on a t-distribution with an Upper Bound for the Degrees of Freedom.

Covariate	*MCMC mean*	*HPD95 lower*	*HPD95 upper*	*pMCMC*	*Pr(>|t|)*
(Mean Age, Average Reader, None, Random)	-0.578	-1.210	0.059	0.0759	0.0957
Age	0.024	-0.037	0.084	0.4332	0.4440
Dyslexic Reader	-0.006	-0.884	0.848	0.9875	0.9958
Hysteresis	8.347	7.693	9.027	0.0001	0.0000
Enhanced Contrast	-6.941	-7.433	-6.486	0.0001	0.0000
B>D>B	0.480	-0.134	1.101	0.1249	0.1206
D>B>D	1.103	0.497	1.743	0.0004	0.0005
Age:Dyslexic Reader	0.001	-0.063	0.080	0.7747	0.8133
Age:Hysteresis	-0.032	-0.095	0.025	0.2952	0.2880
Age:Enhanced Contrast	-0.051	-0.101	-0.002	0.0439	0.0415
Dyslexic Reader:Hysteresis	-0.309	-1.215	0.605	0.5007	0.5064
Dyslexic Reader:Enhanced Contrast	0.133	-0.466	0.756	0.6628	0.6863
Age:B>D>B	-0.053	-0.113	0.004	0.0735	0.0717
Age:D>B>D	-0.056	-0.115	-0.002	0.0545	0.0496
Dyslexic Reader:B>D>B	0.081	-0.811	0.891	0.8569	0.8612
Dyslexic Reader:D>B>D	-0.108	-0.941	0.702	0.8085	0.7945
Hysteresis:B>D>B	-4.103	-4.983	-3.240	0.0001	0.0000
Enhanced Contrast:B>D>B	-0.572	-1.224	0.078	0.0863	0.0802
Hysteresis:D>B>D	-4.925	-5.789	-4.091	0.0001	0.0000
Enhanced Contrast:D>B>D	1.910	1.315	2.638	0.0001	0.0000
Age:Dyslexic Reader:Hysteresis	0.008	-0.064	0.079	0.8132	0.7963
Age:Dyslexic Reader:Enhanced Contrast	-0.028	-0.085	0.025	0.3188	0.3290
Age:Dyslexic Reader:B>D>B	0.093	0.003	0.174	0.0317	0.0315
Age:Dyslexic Reader:D>B>D	0.066	-0.014	0.140	0.0912	0.0953
Age:Hysteresis:B>D>B	-0.133	-0.203	-0.061	0.0001	0.0003
Age:Enhanced Contrast:B>D>B	0.218	0.156	0.279	0.0001	0.0000
Age:Hysteresis:D>B>D	-0.094	-0.163	-0.024	0.0073	0.0083
Age:Enhanced Contrast:D>B>D	0.136	0.077	0.196	0.0001	0.0000
Dyslexic Reader:Hysteresis:B>D>B	3.054	1.835	4.326	0.0001	0.0000
Dyslexic Reader:Enhanced Contrast:B>D>B	-1.171	-2.091	-0.304	0.0103	0.0113
Dyslexic Reader:Hysteresis:D>B>D	1.386	0.146	2.527	0.0256	0.0232
Dyslexic Reader:Enhanced Contrast:D>B>D	-0.712	-1.579	0.166	0.1096	0.1131
Age:Dyslexic Reader:Hysteresis:B>D>B	-0.087	-0.190	0.017	0.0945	0.0885
Age:Dyslexic Reader:Enhanced Contrast:B>D>B	-0.160	-0.246	-0.070	0.0009	0.0004
Age:Dyslexic Reader:Hysteresis:D>B>D	0.029	-0.067	0.123	0.5644	0.5771
Age:Dyslexic Reader:Enhanced Contrast:D>B>D	-0.145	-0.225	-0.064	0.0004	0.0005

Note: Number of observations = 44640, Groups: participant ID = 186; Stimulus number = 20

Notes

Summary [Dutch]

Een korte samenvatting in het Nederlands.

De Begrenzing Voorbij.

Een analyse van verisimilitude en causale ontologie van wetenschappelijke claims:
De etiologie van ontwikkelingsdyslexie als klinkend voorbeeld.

Stel dat bij uw kind de diagnose ‘ontwikkelingsdyslexie’ is vastgesteld. Als ouder wilt u natuurlijk de beste behandeling voor de leesproblemen van uw kind en daarom raadpleegt u de recente wetenschappelijke literatuur en selecteert alleen studies waarin de effectiviteit van een behandeling wordt onderzocht, en ook aangetoond. Een greep uit de effectieve behandelingen die u zult tegenkomen:

2001 Gekleurde lenzen en transparante overlays gebruiken om de visuele stress die uw kind ervaart te verminderen.

2004 Intensieve training met toon en gemodificeerde spraak om de beperking in de temporele auditieve verwerking van uw kind te verhelpen.

2005 Visuele-, auditieve- en tastzintuigen stimuleren om de disfunctionele hersenhelft van uw kind te te remediëren.

2007 Visolie (Omega--3 vetzuren) toevoegen aan het dieet om het tekort in de myelinisatie van de zenuwcellen van uw kind op te heffen.
Trainen van motorische vaardigheden om de motorische beperkingen van uw kind te verhelpen.

2012 Muziek therapie, om de de beperking in ritme-perceptie van uw kind te behandelen.

2013 Laat uw kind actie-computerspelletjes spelen om zo het tekort in de visuo-spatiele, cross-modale, temporele, aandacht op te heffen.

De merkwaardige situatie doet zich voor dat al deze behandelingen tegelijkertijd ‘werkzaam’ zijn, waardoor de verklaringen voor hun werkzaamheid ook tegelijkertijd ‘waar’ zijn. De behandelingen zijn gebaseerd op wetenschappelijke claims over de oorzaken van ontwikkelingsdyslexie die nog altijd leiden tot nieuw, publicabel empirisch onderzoek.

Zou het niet zo moeten zijn dat het herhaaldelijk toepassen van de wetenschappelijke methode, als ultieme test voor theoretische claims over hoe de werkelijkheid in elkaar zit, moeten leiden tot een beperkt aantal theorieën die elkaar beconcurreren op hun *waar*achtigheid (*veri*similitude)?

Omgaan met Theoretische Diversiteit

In het eerste hoofdstuk van dit proefschrift wordt gesteld dat (i) de veelheid aan theorieën over de oorzaken van dyslexie problematisch is, (ii) vrijwel alle onderzoeksvelden in de sociale en levenswetenschappen met dit probleem te maken hebben, en (iii) dat een mogelijke oorzaak het ‘component-denken’ is.

In het component-denken, of, *de component-dominante causale ontologie* waarin het doel van het verklaren van (intelligent) gedrag gereduceerd wordt tot het vinden van unieke oorzaken, doorgaans geïdentificeerd op een 'lagere', of meer bio-fysische schaal van observatie (neuronen, genen, aangeboren vaardigheden). De tegenhanger is het 'interactie-denken', of, de *interactie-dominante causale ontologie* waarin het niet de componenten van een systeem zijn die (intelligent) gedrag veroorzaken, maar juist de interacties tussen die componenten. Het aanwijzen van unieke oorzaken is volgens deze causale ontologie niet mogelijk, of zinvol.

Ter illustratie, de methoden en technieken die bij de component-dominante ontologie horen zijn de lineaire statistiek en factor-designs, die latente bronnen van variantie proberen aan te tonen. De technieken die bij de interactie-dominante ontologie horen zijn gebaseerd op de analyse van de niet-lineaire dynamica van complex systemen zoals perturbatie analyse, schaal-invartiantie van observabelen en de sensitieve afhankelijkheid van initiële condities.

Na uitgebreide analyse van mogelijke oorzaken en gevolgen van het component-denken wordt gesuggereerd dat een interactie-dominante ontologie mogelijk een meer coherente verklaring van een groot aantal gedragsfenomenen zou kunnen geven. Er worden een aantal principes opgesteld die algemeen toepasbaar zouden kunnen zijn om theorieën die binnen de sociale en levenswetenschappen geponeerd worden formeler te kunnen evalueren. Deze principes worden toegepast op verklaringen voor ontwikkelingsdyslexie die gebaseerd zijn op een deficiet in de spraakperceptie. Hoofdstuk 2 tot en met 5 doen verslag van resultaten en interpretatie van empirische studies naar spraakperceptie bij dyslectische lezers, vanuit het perspectief van de interactie-dominante causale ontologie

Hoofdstuk 2 - Context relativiteit van orde relaties en prospectieve predictie

In dit hoofdstuk wordt onderzocht of de individuele variabiliteit in het waarnemen van spraak (beslissingssnelheid), van kleuters met een genetisch risico op dyslexie gerelateerd is aan hun leesvaardigheid 1 jaar later. De resultaten laten zien dat kinderen die tot 25% laagste lezers behoren inderdaad in de kleuterleeftijd meer variabiliteit in hun reactietijden lieten zien.

Hoofdstuk 3 - Principiële simulatie van orde relaties.

De resultaten uit Hoofdstuk 2 worden gesimuleerd met een recurrent neuraal netwerk model waarbij wordt verondersteld dat de dyslectische lezers een instabieler netwerk hebben dan de gemiddelde lezer. De orde relaties uit Hoofdstuk 2 worden conceptueel gerepliceerd door de neurale netwerken.

Hoofdstuk 4 - Sterke Inferentie

In hoofdstuk 4 worden zogenaamde 'condities voor sterke inferentie' gecreëerd waardoor een directe evaluatie van drie hypothesen over eigenschappen van het spraaksignaal die essentiële zijn voor de spraakperceptie kan plaatsvinden. De hypothesen nemen ieder een plaats in op het spectrum van component-dominante en interactie-dominante ontologie. De analyses laten zien dat een classificatie algoritme de beoordelingen van zowel dyslectische als niet-dyslectische lezers het beste kan repliceren op basis van de fysische eigenschappen van het spraaksignaal die aangegeven worden door de interactie-dominante ontologie.

Hoofdstuk 5 - Principiële simulatie van theoretische constructen en Sterke Inferentie

In hoofdstuk 5 worden op basis van de literatuur over spraakperceptie en ontwikkelingsdyslexie, en potentiaal theorie (niet-lineaire dynamica van attractoren van an een fysisch systeem) drie voorspellingen getoetst die bij bevestiging een serieus probleem voor de evaluatie van de verisimilitude van component theorieën. De hypothesen betreffen de discreet veronderstelde, categorische aard van de klankrepresentatie, de uniformiteit van de interne structuur van een klankcategorie en de biologische ontwikkeling van deze categorieën, gebaseerd op een zogenaamde 'koppelingshypothese'. Op basis van tweedimensionaal potentiaal model dat een nadere uitwerking is van het neurale netwerkmodel uit Hoofdstuk 3 wordt geconcludeerd dat de component-dominante ontologie voor de verklaring van mogelijke verschillen in de spraakperceptie tussen dyslectische en niet-dyslectische lezers onhoudbaar is.

In Hoofdstuk 6 wordt een algemene interpretatie gegeven aan de resultaten en worden een aantal lijnen uitgezet naar toekomstig theoretisch en empirisch onderzoek.

Notes

How Did We Do? [So Far]

ALL IN A DAY'S DREAM:
AN EPILOGICAL PROLOGUE[a]

The Master of the Labyrinth sighed deeply as he knelt beside me.
I saw his eyes focus on my naked brain and I guessed he was examining it for abnormalities.
He then spoke to me:

Foolish young creature, your quest continues.
I will share with you the Truths you must pursue,
for they are One and they are All. First, analyse
the manifestations of Order, then delve for the
powers of Chaos. When you are still in possession
of your life after having done these things,
you must begin the search for the Lost Archetype.
Be everywhere, investigate nowhere.

Then he took a piece of my brain
and as he slowly merged with the dark surroundings
he said: ``**You won't be needing this.**''

-Fred Hasselman, 1992 (Song Lyrics)

[a]To the songs: *Confessions of Demi-God Parts 1&2: Random Order & Sentient Chaos*

Acknowledgements, Odes and Reveries

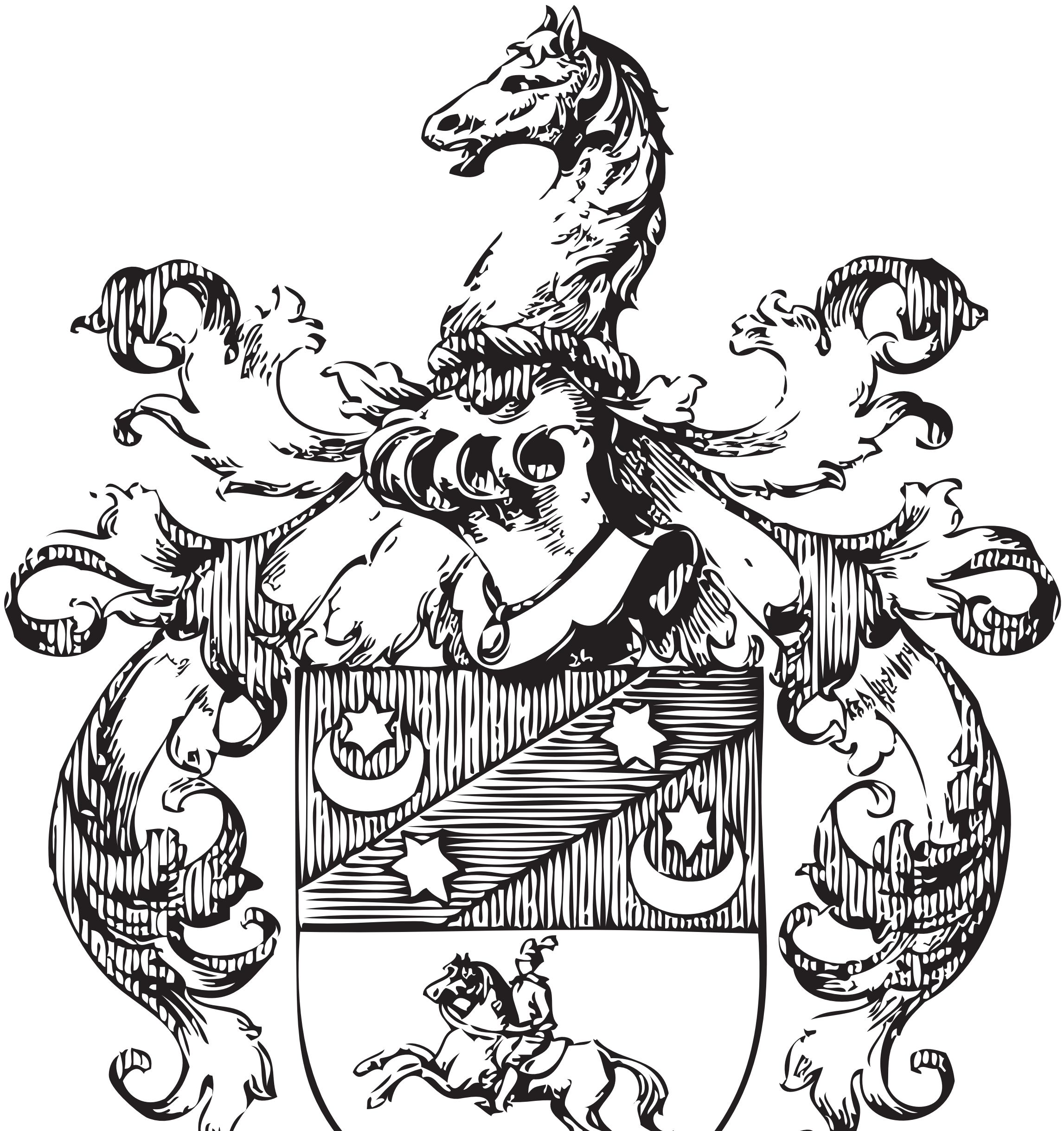
Beginnings are very important. I owe a lot to experience gained working as a 'Research Assistant' for the RTD and RTOG and as a 'Junior Docent' for the schools of Psychology and Artificial Intelligence and School of Pedagogical and Educational Science.

In-the-meantime (in-between start and finish), is a state I seem to prefer. There are many colleagues without whom this book would not have reached its finished state, most notably: Anna Bosman, Guy Van Orden, various incarnations of the Dynamic/Complex System group, Ralf Cox, and Ludo Verhoeven.

The journey from possibility to realization is long and hard and is not endured in solitude. I owe the world to the support of my mother, brother and sister, and my friends who never doubted me, and Noëlla and Robin, the loves of my life.

Endings are a fact of life. I deeply regret the Goddess called these men to join her side, before I was able to discuss the contents of this book with them:

My Father
Guy Van Orden
Jan van Leeuwe
Rob Schreuder
Michel Ekkebus



Vitæ[Short]

The author was born in Geldrop on the 25th of December, 1973; raised in Aalst/Waalre; went to High-School in Eindhoven, and to University in Nijmegen, where he studied Cognitive Psychology [Funktieleer].

The origins of this dissertation lie with a PhD project that started in 2001. In the meantime, he held jobs as a statistical advisor and junior and senior lecturer at the Faculty of Social Science of the Radboud University.

Currently employed by the School of Pedagogical and Educational Science, he teaches Research Methods and Statistics in the 1st Bachelor year. In the Research Master Behavioural Science, he teaches the course Dynamics of Complex Systems.

The author is an active member of the Open Science Collaboration (and affiliated initiatives) for improving openness and transparency in Science. For example, the *Reproducibility Project: Psychology* (https://osf.io/ezcuj/wiki/home/) and *Manylabs2* (https://osf.io/8cd4r/).

His research interests concern the study of intelligent behavior based on principles of biology and physics, which opens up the broad range of topics in Complexity Science, from philosophy of the embodied mind to the assignment of properties through measurement of system observables that are inherently complex, dynamical and due to interactions on multiple-scales.

The author published about 30 papers, which received about 350 citations.
Google Scholar says his *H-Index = 10*.
Scopus Author ID says his *H-Index = 9*.
ResearcherID says his *H-Index = 7*.

A list of recent publications and citation information can be found here:

Google Scholar: https://scholar.google.com/citations?user=yIlTr6QAAAAJhl=en
ORCID: http://orcid.org/0000-0003-1384-8361
Scopus Author ID: http://www.scopus.com/authid/detail.url?authorId=24278548700
ResearcherID: http://www.researcherid.com/rid/C-5603-2011

Personal Site: http://fredhasselman.com

Bibliography

Abney DH, Paxton A, Kello CT, Dale R. 2005. Complexity Matching in Dyadic Interaction. Journal of Experimental Psychology. General 143:2304–2315.

Abramson, A. (1979). The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Öohman (Eds.), Frontiers of speech communication research (pp. 127–134). London: Academic Press. Retrieved from http://web.haskins.yale.edu/Reprints/HL0279.pdf

Abramson, A., & Lisker, L. (1970). Discriminability along the voicing continuum: cross-language tests. Proceedings of the sixth international congress of phonetic sciences (pp. 569–573). Academia Prague, Czechoslovakia.

Abramson, A., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. Journal of Phonetics, 1, 1–8. Retrieved from http://www.haskins.yale.edu/Reprints/HL0128.pdf

Ackermann, H., Mathiak, K., & Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: clinical and functional imaging data. Cerebellum, 6(3), 202–13. doi:10.1080/14734220701266742

Adolph, K. E. (1997). The Growing Body In Action: What Infant Locomotion Tells Us About Perceptually Guided Action. psych.nyu.edu, (1987). Retrieved from http://www.psych.nyu.edu/adolph/PDFs/AdolphCMUChap.pdf

Adolph, K. E. (2008). Learning to Move. Current directions in Psychological Science, 17(3), 213–218. doi:10.1111/j.1467-8721.2008.00577.x

Aihara T, Kitajo K, Nozaki D, Yamamoto Y. 2010. How does stochastic resonance work within the human brain? - Psychophysics of internal and external noise. Chemical Physics 375:616–624.

Ainsworth L, Vela C. 2013. Letters. Nursing 43:8.

Akhromeeva, T. S., & Malinetskii, G. G. (2009). Synergetics and measurement problems. Measurement Techniques, 51(11), 1155–1161. doi:10.1007/s11018-009-9180-z

Aoyama, T., Hayakawa, M., Kinoshita, T., & Nio, M. (2008). Revised value of the eighth-order QED contribution to the anomalous magnetic moment of the electron. Physical Review D, 77(5), 1–24. doi:10.1103/PhysRevD.77.053012

Aquino G, Bologna M, West BJ, Grigolini P. 2011. Transmission of information between complex systems: 1/f resonance. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 83:051130.

Aßmann B, Romano MC, Thiel M, Niemitz C. 2007. Hierarchical organization of a reference system in newborn spontaneous movements. Infant Behavior and Development 30:568–586.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59(4), 390–412. doi:10.1016/j.jml.2007.12.005

Bakan, D. (1966). The test of significance in psychological research. Psychological Bulletin, 66(6), 423–37.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The Rules of the Game Called Psychological Science. Perspectives on Psychological Science, 7(6), 543–554. doi:10.1177/1745691612459060

Balsters, J. H., Cussans, E., Diedrichsen, J., Phillips, K. a, Preuss, T. M., Rilling, J. K., & Ramnani, N. (2010). Evolution of the cerebellar cortex: the selective expansion of prefrontal-projecting cerebellar lobules. NeuroImage, 49(3), 2045–52. doi:10.1016/j.neuroimage.2009.10.045

Barrat, A, Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America, 101(11), 3747–52. doi:10.1073/pnas.0400087101

Barrett NF. 2011. Radical embodied cognitive science. The MIT Press.

Barrett, J., & Kent, A. (2004). Non-contextuality, finite precision measurement and the Kochen–Specker theorem. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 35(2), 151–176. doi:10.1016/j.shpsb.2003.10.003

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. Developmental Psychobiology, 40(3), 293–310. doi:10.1002/dev.10034

Bates, E., Thal, D., Finlay, B., & Clancy, B. (2002). Early language development and its neural correlates. In S. J. Segalowitz & I. Rapin (Eds.), Handbook of Neuropsychology, Vol. 8. Part II (2nd ed., Vol. 8). Amsterdam: Elsevier Science

Been, P. H., & Zwarts, F. (2003). Developmental dyslexia and discrimination in speech perception: a dynamic model study. Brain and Language, 86, 395-412.

Been, P. H., & Zwarts, F. (2004). Language disorders across modalities: The case of developmental dyslexia. In H. van Balkom & L. Verhoeven (Eds.), Classification of developmental language disorders: Theoretical issues and clinical implications (pp. 61-97). Mahwah, NY: Lawrence Erlbaum Associates Publishers.

Been, P. H., van Leeuwen, T. H., van Herten, M., Maassen, B., van der Leij, A., & Zwarts, F. (2008). Auditory P2 is reduced in 5 month old infants from dyslexic families. In Z. Breznitz (Ed.), Brain Research in Language (Literacy Studies Vol. 1, pp. 93–110). Springer US. doi:10.1007/978-0-387-74980-8_3

Beinum, F. J. Van, Schwippert, C. E., Been, P. H., Leeuwen, T. H. Van, Kuijpers, C. T. L. L., van Beinum, F. J., & van Leeuwen, T. H. (2005). Development and application of a /bAk/–/dAk/ continuum for testing auditory perception within the Dutch longitudinal dyslexia study. Speech Communication, 47(1-2), 124–142. doi:10.1016/j.specom.2005.04.003

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. Journal of Personality and Social Psychology, 407–425. doi:10.1037/a0021524

Bennett, G., Bousquet, B., Brown, H., Bunce, G., Carey, R., Cushman, P., … Zimmerman, D. (2006). Final report of the E821 muon anomalous magnetic moment measurement at BNL. Physical Review D, 73(7), 1–41. doi:10.1103/PhysRevD.73.072003

Bianciardi M, Sirabella P, Hagberg GE, Giuliani A, Zbilut JP, Colosimo A. 2007. Model-free analysis of brain fMRI data by recurrence quantification. NeuroImage 37:489–503.

Blau, V., Reithler, J., van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., & Blomert, L. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. Brain : A Journal of Neurology, 133(Pt 3), 868–79. doi:10.1093/brain/awp308

Blomert, L. (2005). Dyslexie in Nederland. Amsterdam: Uitgeverij Nieuwezijds. Retrieved from http://www.nieuwezijds.nl/boeken/BladerPDF/9789057122170.pdf

Blomert, L. (2010). The neural signature of orthographic-phonological binding in successful and failing reading development. NeuroImage. doi:10.1016/j.neuroimage.2010.11.003

Blomert, L., & Mitterer, H. (2004). The fragile nature of the speech-perception deficit in dyslexia: Natural vs. synthetic speech. Brain and Language, 89(1), 21–26. doi:10.1016/S0093-934X(03)00305-5

Blomert, L., & Vaessen, A. (2009). 3DM differential diagnostics for dyslexia: cognitive analysis of reading and spelling. Amsterdam: Boom Test Publishers.

Blomert, L., Mitterer, H., & Paffen, C. (2004). In Search of the Auditory, Phonetic, and/or Phonological Problems in Dyslexia: Context Effects in Speech Perception. Journal of Speech, Language, and Hearing Research, 47(5), 1030.

Boersma P, Weenink D. (2002). Praat: Doing phonetics by computer (Version 4.0) [Computer Program].

Boets B, Ghesquière P, van Wieringen A, Wouters J. 2007. Speech perception in preschoolers at family risk for dyslexia: Relations with low-level auditory processing and phonological ability. Brain and Language 101:19–30.

Bogliotti, C., Serniclaes, W., Messaoud-Galusi, S., & Sprenger-Charolles, L. (2008). Discrimination of speech sounds by children with dyslexia: comparisons with chronological age and reading level controls. Journal of Experimental Child Psychology, 101(2), 137–55. doi:10.1016/j.jecp.2008.03.006

Bokulich, A. (2004). Open or Closed? Dirac , Heisenberg , and the Relation between, 35, 1–31.

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient to Rule Out Placebo Effects. Perspectives on Psychological Science, 8(4), 445–454. doi:10.1177/1745691613491271

Born, M., Heisenberg, W., & Jordan, P. (1926). Zur Quantenmechanik. II. Zeitschrift für Physik, 35(8-9), 557–615.

Borsboom D. (november, 20, 2013). Theoretical Amnseia. http://osc.centerforopenscience. org/2013/11/20/theoretical-amnesia/

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. PloS One, 6(11), e27407. doi:10.1371/journal.pone.0027407

Bosman, A. M. T., & De Groot, A. M. B. (1995). Evidence for assembled phonology in beginning and fluent readers as assessed with the first-letter-naming task. Journal of Experimental Child Psychology, 59, 234–259.

Bosman, A. M. T., & De Groot, A. M. B. (1996). Phonologic mediation is fundamental to reading: Evidence from beginning readers. The Quarterly Journal of Experimental Psychology Section A, 49(3), 715–744. doi:10.1080/713755641

Bosman, A. M. T., Cox, R. C. A., Hasselman, F., & Wijnants, M. L. (2013). From the Role of Context to the Measurement Problem: The Dutch Connection Pays Tribute to Guy Van Orden. Ecological Psychology, 25(3), 240–247.

Bosman, A. M. T., Van Leerdam, M., & De Gelder, B. (2000). The /O/ in OVER is different from the /O/ in OTTER: Phonological effects in children with and without dyslexia. Developmental Psychology, 36(6), 817–825. doi: 10.1037//0012-1649.36.6.817

Bower, B. (1997). Null Science. Science News, 151(23).

Boyle C. (1994). Computation as an intrinsic property. Minds and Machines 15237:451–467.

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. Nature, 271(5647), 746–747. doi:10.1038/271746a0

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—a causal connection. Nature, 301(5899), 419–421. doi:10.1038/301419a0

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. a., Perugini, M., Spies, J. R., and van 't Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? Journal of Experimental Social Psychology, 50:217–224.

Bravi A, Longtin A, Seely AJ. 2011. Review and classification of variability analysis techniques with clinical applications. BioMedical Engineering OnLine 10:90.

Breteler, M. H. M., Arns, M., Peters, S., Giepmans, I., & Verhoeven, L. (2010). Improvements in spelling after QEEG-based neurofeedback in dyslexia: a randomized controlled treatment study. Applied Psychophysiology and Biofeedback, 35(1), 5–11. doi:10.1007/s10484-009-9105-2

Brookes, R. L., Tinkler, S., Nicolson, R. I., & Fawcett, A. J. (2010). Striking the Right Balance: Motor Difficulties in Children and Adults with Dyslexia. Dyslexia, 373, 358–373. doi:10.1002/dys

Brown, H. R., & Uffink, J. (2001). The Origins of Time-Asymmetry in Thermodynamics: The Minus First Law. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 32(4), 525–538. doi: 10.1016/S1355-2198(01)00021-1

Browne WJ, Steele F, Golalizadeh M, Green MJ. 2009. The use of simple reparameterizations to improve the efficiency of Markov chain Monte Carlo estimation for multilevel models with applications to discrete time survival models. Journal of the Royal Statistical Society. Series A: Statistics in Society 172:579–598.

Browne, W. J. (2009). MCMC Estimation in MLwiN v2.1. Bristol: Centre for Multilevel Modelling, University of Bristol.

Brush, S. (1989). Prediction and Theory Evaluation: The Case of Light Bending. Science, 246(4934), 1124–1129.

Bunge, M. (1967). Physical Axiomatics. Reviews of Modern Physics, 39(2), 463–474.

Bunge, M. (1968). Physical Time: The Objective and Relational Theory. Philosophy of Science, 35(4), 355–388. doi: 10.2307/186253

Busemeyer, J. r., Pothos, E. M., Franco, r., & Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. Psychological review, 118(2), 193–218. doi:10.1037/a0022542

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. a., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience, 14(May). doi:10.1038/nrn3475

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. Journal of speech, language, and hearing research: JSLHR, 43(3), 721–36.

Camaioni, L., Aureli, T., Bellagamba, F., & Fogel, A. (2003). a longitudinal examination of the Transition to Symbolic Communication in the Second Year of Life. Infant and Child Development, 12(1), 1–26. doi:10.1002/icd.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. The Journal of the Acoustical Society of America, 62(4), 961.

Carver, R. (1993). The Case Against Statistical Significance Testing, Revisited. The Journal of Experimental Educational, 61, 287–292. doi: 10.2307/20152382

Case, P., Tuller, B., Ding, M., & Kelso, J. (1995). Evaluation of a dynamical model of speech perception. Perception & Psychophysics, 57(7), 977–88.

Caylak, E. (2009). Neurobiological approaches on brains of children with dyslexia: review. Academic Radiology, 16(8), 1003–24. doi:10.1016/j.acra.2009.02.012

Cesario, J. (2014). Priming, Replication, and the Hardest Science. Perspectives on Psycho- logical Science, 9(1):40–48.

Chemero A, Turvey MT. 2007. Complexity, Hypersets, and the Ecological Perspective on Perception-Action. Biological Theory 2:23–36.

Chemero, A., & Turvey, M.T. (2010). Is Life Computable? Advances in Modeling Adaptive and Cognitive Systems, 29–37.

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., et al. (1995). Mismatch negativity indicates vowel discrimination in newborns. Hearing Research, 82(1), 53–8.

Chomsky, N., & Halle, M. (1968). The Sound Pattern of English. Harper & Row.

Clark, A. (1999). An embodied cognitive science? Trends in Cognitive Sciences, 3(9), 345–351. doi:10461197

Cochrane, A. (1972). Effectiveness and Efficiency: Random reflections on health services. Oxfordshire, GB: The Nuffield Provincial hospitals Trust. Retrieved from http://pgim.srilankahealthrepository.org/handle/123456789/9143

Coey CA, Washburn A, Richardson MJ. 2014. Translational Recurrences. In: Marwan N, Riley M, Giuliani A, Webber, CL eds. Translational Recurrences. Springer Proceedings in Mathematics & Statistics. Cham: Springer International Publishing, 1–14.

Cogo-Moreira, H., Andriolo, R. B., Yazigi, L., Ploubidis, G. B., Brandão de Ávila, C. R., & Mari, J. J. (2012). Music education for improving reading skills in children and adolescents with dyslexia. The Cochrane Database of Systematic Reviews, 8(8), CD009133. doi:10.1002/14651858.CD009133.pub2

Cohen, J. (1994). The earth is round (p<. 05). American Psychologist, 49(12), 997–1003.

Colangelo, A., Holden, J. G., Buchanan, L., & Van Orden, G. C. (2004). Speculation about behavior, brain damage, and self-organization: The other way to herd a cat. Brain and Language, 90(1-3), 151–159. doi:10.1016/S0093-934X(03)00428-0

Collins JJ, Chow CC, Imhoff TT. 1995. Stochastic resonance without tuning. Nature 376:236–238.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some Experiments on the Perception of Synthetic Speech Sounds. Journal of the Acoustical Society of America, (3802), 30–36.

Correll, J. (2008). 1/F Noise and Effort on Implicit Measures of Bias. Journal of Personality and Social Psychology, 94(1), 48–59. doi:10.1037/0022-3514.94.1.48

Correll, J. (2011). Order from chaos? 1/f Noise predicts performance on reaction time measures. Journal of Experimental Social Psychology, 47(4), 1–6. doi:10.1016/j.jesp.2011.02.019

Corriveau K, Pasquini E, Goswami U. 2007. Basic auditory processing skills and specific language impairment: a new look at an old hypothesis. Journal of speech, language, and hearing research: JSLHR 50:647–666.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 45(1), 119–30. doi: 10.1016/j.cortex.2007.09.008

Cox, R. F. A., & Hasselman, F. (2013). The Case of Watson vs. James: Effect-Priming Studies Do Not Support Ideomotor Theory. PLoS ONE, 8(1), e54094. doi:10.1371/journal.pone.0054094

Cox, R. F. A., Hasselman, F., & Seevinck, M. P. (2011). A Process Approach to Psychological Measurement: Lessons from Quantum Mechanics. Proceedings of the 16th International Conference on Perception-Action.

Cronbach, L., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 59(4), 257–72.

Cumming, G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis (pp. 1–9). Routledge New York.

Cyhlarova, E., Bell, J. G., Dick, J. R., Mackinlay, E. E., Stein, J. F., & Richardson, A. J. (2007). Membrane fatty acids, reading and spelling in dyslexic and non-dyslexic adults. European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology, 17(2), 116–21. doi:10.1016/j.euroneuro.2006.07.003

Dale R, Spivey MJ. 2006. Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. Language Learning 56:391–430.

Dale, R. (2008). Response: Sloughing ontology. Journal of Experimental & Theoretical Artificial Intelligence, 20(3), 251–256. doi:http://dx.doi.org/10.1080/09528130802319292

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and roc curves. Proceedings of the 23rd international conference on Machine learning (pp. 233–240). New York, New York, USA: ACM. doi: 10.1145/1143844.1143874

De Graaff, S., Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of Systematic Phonics Instruction. Scientific Studies of Reading, 13(4), 318–333. doi:10.1080/10888430903001308

De Graaff, S., Hasselman, F., Bosman, A. M. T., & Verhoeven, L. (2008). Cognitive and linguistic constraints on phoneme isolation in Dutch kindergarteners. Learning and Instruction, 18(4), 391–403. doi:10.1016/j.learninstruc.2007.08.001

De Graaff, S., Hasselman, F., Verhoeven, L., & Bosman, A. M. T. T. (2011). Phonemic awareness in Dutch kindergartners: Effects of task, phoneme position, and phoneme class. Learning and Instruction, 21(1), 163–173. doi:10.1016/j.learninstruc.2010.02.001

De Graaff, S., Verhoeven, L., Bosman, A. M. T., & Hasselman, F. (2007). Integrated pictorial mnemonics and stimulus fading: Teaching kindergartners letter sounds. The British Journal of Educational Psychology, 77(Pt 3), 519–39. doi:10.1348/000709906X160011

De Graag, J. A., Cox, R. F. A., Hasselman, F., Jansen, J., & De Weerth, C. (2012). Functioning within a relationship: Mother-Infant Synchrony and Infant Sleep. Infant Behavior & Development, 35(2), 252–63. doi:10.1016/j.infbeh.2011.12.006

DeCasper, A. J., & Spence, M. J. (1986). Prenatal Maternal Speech Influences Newborns' Perception of Speech Sounds. Infant Behavior and Development, 9(2), 133–150.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. NeuroImage, 24(1), 21–33. doi:10.1016/j.neuroimage.2004.09.039

Dijksterhuis, A. (2013). Replication crisis or crisis in replication? A reinterpretation of Shanks et al. [Peer Commentary on "Priming Intelligent Behavior: An Elusive Phenomenon" by A. Dijksterhuis]. PLoS ONE, 8(4):e56515.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of Cognates and Interlingual Homographs: The Neglected Role of Phonology. Journal of Memory and Language, 41(4), 496–518. doi:10.1006/jmla.1999.2654

Diniz A, Wijnants ML, Torre K, Barreiros J, Crato N, Bosman AMT, Hasselman F, Cox RF a, Van Orden GC, Delignières D. 2011. Contemporary theories of 1/f noise in motor control. Human Movement Science 30:889–905.

Dirac, P. A. M. (1930). The principles of quantum mechanics. Oxford,: The Clarendon Press.

Dmitrieva, L. P., & Gottlieb, G. (1994). Influence of auditory experience on the development of brain stem auditory-evoked potentials in mallard duck embryos and hatchlings. Behavioral and Neural Biology, 61(1), 19–28. doi:10.1016/S0163-1047(05)80040-8

Doupe, a J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. Annual Review of Neuroscience, 22, 567–631. doi:10.1146/annurev.neuro.22.1.567

Dowker, A. (2006). What can functional brain imaging studies tell us about typical and atypical cognitive development in children? Journal of Physiology, Paris, 99(4-6), 333–41. doi:10.1016/j.jphysparis.2006.03.010

Dreyfus, H. L. (2002). Intelligence without representation – Merleau-Ponty ' s critique of mental representation The relevance of phenomenology to scientific explanation. Phenomenology and the Cognitive Sciences, (1), 367–383. doi:10.1023/A:1021351606209

Duca, P. (2005). ROC curve analysis and medical decision making: what's the evidence that matters for evidence-based diagnosis? Proceedings of the 1st European Workshop on the (pp. 33–50). Retrieved from http://www.tech.plym.ac.uk/spmc/pdf/EWADP2004/EWADP2004.PAPER03.P33-49.DUCA.pdf

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? Annual review of psychology, 55, 51–86. doi:10.1146/annurev.psych.55.090902.142050

Durbin, R., Eddy, S., Krogh, A., & Mitchison, G. (1998). Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge, UK: Cambridge University Press.

Eckert, M. (2004). Neuroanatomical markers for dyslexia: a review of dyslexia structural imaging studies. The Neuroscientist: A review journal bringing neurobiology, neurology and psychiatry, 10(4), 362–71. doi:10.1177/1073858404263596

Edelman, G. M. (1987). Neural Darwinism: The theory of neuronal group selection. Basic Books.

Efron B. 2011. The Bootstrap and Markov-Chain Monte Carlo. Journal of Biopharmaceutical Statistics 21:1052–1062.

Eilers, R. E., Gavin, W., & Wilson, W. R. (1979). Linguistic experience and phonemic perception in infancy: a crosslinguistic study. Child Development, 50(1), 14–8.

Eimas, P., & Miller, J. (1980). Discrimination of information for manner of articulation. Infant Behavior and Development, 3, 367–375. doi:10.1016/S0163-6383(80)80044-0

Einstein, A. (1934\1952) Ideas and Opinions (pp. 100-105) (S Bargmann trans.). Crown: New York, NY, USA.

Elbro, C., Borstrøm, I., & Petersen, D. K. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. Reading Research Quarterly, 33(1), 36–60.

Eleveld, M. (2005). At risk for dyslexia: the role of phonological abilities, letter knowledge and speed of serial naming in early intervention and diagnosis. University of Groningen.

Elliott, J. G., & Gibbs, S. (2008). Does Dyslexia Exist? Journal of Philosophy of Education, 42(3-4), 475–491. doi:10.1111/j.1467-9752.2008.00653.x

F. Takens, 1981. Detecting strange attractors in fluid turbulence. In: Rand D, Young L-S eds. Dynamical Systems and Turbulence. Lecture Notes in Mathematics. Springer Berlin / Heidelberg, 366–381.

Fajen BR, Matthis JS. 2011. Direct perception of action-scaled affordances: The shrinking gap problem. Prentice-Hall Englewood Cliffs, NJ.

Fanelli, D. (2010). "Positive" results increase down the Hierarchy of the Sciences. PloS One, 5(4), e10068. doi:10.1371/journal.pone.0010068

Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. Psychonomic Bulletin & Review, 2(4), 460–493.

Farnsworth, H. (2001). Roswell that Ends Well. Futurama.

Farrar, W. T., & Van Orden, G. C. (2001). Errors as multistable response options. Nonlinear Dynamics, Psychology, and Life Sciences, 5(3), 223–265.

Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. Machine Learning, 31, 1–38.

Feldman M. 2008. Theoretical analysis and comparison of the Hilbert transform decomposition methods. Mechanical Systems and Signal Processing 22:509–519.

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null . Perspectives on Psychological Science , 7 (6), 555–561. doi:10.1177/1745691612459059

Fiedler K, Kutzner F, Krueger J. 2012. The Long Way From α-Error Control to Validity Proper Problems With a Short-Sighted False-Positive Debate. Perspectives on Psychological Science 7:661–669.

Fisher, R. (1958). The Nature of Pobability. Centennial Review, 2, 261–274. Retrieved from http://www.york.ac.uk/depts/maths/histstat/fisher272.pdf

Fletcher, J. M. (2009). Dyslexia: The evolution of a scientific concept. Journal of the International Neuropsychological Society: JINS, 15(4), 501–8. doi:10.1017/S1355617709090900

Fletcher, J. M., & Lyon, G. R. (2008). Dyslexia: Why Precise Definitions Are Important and How We Have Achieved Them. Language, 27–31.

Forgeard, M., Winner, E., Norton, A., & Schlaug, G. (2008). Practicing a musical instrument in childhood is associated with enhanced verbal ability and nonverbal reasoning. PloS One, 3(10), e3566. doi:10.1371/journal.pone.0003566

Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M., & Facoetti, A. (2013). Action video games make dyslexic children read better. Current Biology: CB, 23(6), 462–6. doi:10.1016/j.cub.2013.01.044

Freeman WJ, Kozma R, Werbos PJ. 2001. Biocomplexity: Adaptive behavior in complex stochastic dynamical systems. BioSystems 59:109–123.

Freeman, W. J. (1999). Consciousness, intentionality and causality. Journal of Consciousness Studies, 6(11-12), 143–172.

Galaburda, A. M., & Kemper, T. L. (1979). Cytoarchitectonic abnormalities in developmental dyslexia: a case study. Annals of Neurology, 6(2), 94–100. doi:10.1002/ana.410060203

Galaburda, A. M., LoTurco, J., Ramus, F., Fitch, R. H., & Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. Nature Neuroscience, 9(10), 1213–7. doi:10.1038/nn1772

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. Psychonomic Bulletin & Review, 13(3), 361. doi:10.2964/jsik.kuni0223

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. Perception & psychophysics, 66(3), 363–76.

Gibson J.J. (1979). The ecological approach to visual perception. Boston, MA: Houghton, Mifflin and Company; US.

Gibson, J. J. (1966). The Problem of Temporal Order in Stimulation and Perception. The Journal of Psychology, 62(2), 141–149. doi:10.1080/00223980.1966.10543777

Glass, L. (2001). Synchronization and rhythmic processes in physiology. Nature, 410(6825), 277–284.

Gliner, J. A., Vaske, J. J., & Morgan, G. a. (2001). Null Hypothesis Significance Testing: Effect Size Matters. Human Dimensions of Wildlife, 6(4), 291–301. doi:10.1080/108712001753473966

Goldberger AL, Amaral L a N, Hausdorff JM, Ivanov PC, Peng C-K, Stanley HE. 2002. Fractal dynamics in physiology: alterations with disease and aging. Proceedings of the National Academy of Sciences of the United States of America 99 Suppl 1:2466–2472.

Goldstein, L., Nam, H., Saltzman, E. L., & Chitoran, I. (2008). Coupled oscillator planning model of speech timing and syllable structure. Proceedings of the 8th phonetics conference of China and the international symposium on phonetic frontiers.

Goni, J. (2007). Information Theory Toolbox. Pamplona, Spain: University of Navarra - Dpt. of Physics and Applied Mathematics & Centre for Applied Medical Research.

Good, I. J. (1965). The estimation of probabilities: An essay on modern Bayesian methods (Vol. 30). MIT press Cambridge, MA.

Goswami U, Fosker T, Huss M, Mead N, SzuCs D. 2011. Rise time and formant transition duration in the discrimination of speech sounds: The Ba-Wa distinction in developmental dyslexia. Developmental Science 14:34–43.

Goswami U, Gerson D, Astruc L. 2010. Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. Reading and Writing 23:995–1019.

Goswami U, Thomson J, Richardson U, Stainthorp R, Hughes D, Rosen S, Scott SK. 2002. Amplitude envelope onsets and developmental dyslexia: A new hypothesis. Proceedings of the National Academy of Sciences of the United States of America 99:10911–10916.

Goswami, U. (2002). Phonology, reading development, and dyslexia: A cross-linguistic perspective. Annals of Dyslexia, 52(1), 139–163. doi:10.1007/s11881-002-0010-0

Goswami, U. (2006). Sensorimotor impairments in dyslexia: getting the beat. Developmental Science, 9(3), 257–259. doi: 10.1111/j.1467-7687.2006.00485.x

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. Trends in Cognitive Sciences, 15(1), 3–10. doi:10.1016/j.tics.2010.10.001

Goswami, U., Fosker, T., Huss, M., Mead, N., & Szűcs, D. (2010). Rise time and formant transition duration in the discrimination of speech sounds: the Ba-Wa distinction in developmental dyslexia. Developmental Science, 1, 34–43. doi:10.1111/j.1467-7687.2010.00955.x

Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? Applied Psycholinguistics, 24(02), 235–247. doi:10.1017/S0142716403000134

Gottlieb, G. (1987). Development of species identification in ducklings. XIII: A comparison of malleable and critical periods of perceptual development. Developmental Psychobiology, 20(4), 393–404. doi:10.1002/dev.420200404

Gottlieb, G. (1991). Experiential canalization of behavioral development: Theory. Developmental Psychology, 27(1), 4–13. doi:10.1037/0012-1649.27.1.4

Gowen, E., & Miall, R. C. (2007). The cerebellum and motor dysfunction in neuropsychiatric disorders. Cerebellum, 6(3), 268–79. doi:10.1080/14734220601184821

Grasman, R. P. P. P., Van der Maas, H. L. J., & Wagenmakers, E. (2007). Fitting the Cusp Catastrophe in R: A cusp-Package.

Grossberg S. 1999. The link between brain learning, attention, and consciousness. Consciousness and cognition 8:1–44.

Grossberg, S. (2003). Resonant neural dynamics of speech perception. Journal of Phonetics, 31(3-4), 423–445. doi:10.1016/S0095-4470(03)00051-2

Grossberg, S., & Schmajuk, N. (1989). Neural dynamics of adaptive timing and temporal discrimination during associative learning. Neural Networks, 2, 79–102.

Guenther, F. H., & Perkell, J. S. (2004). A neural model of speech production and supporting experiments. In From Sound to Sense Conference. (pp. 98–106). MIT Press. doi:10.1.1.124.1664

Guiard Y. 1993. On Fitts's and Hooke's laws: Simple harmonic movement in upper-limb cyclical aiming. Acta Psychologica 82:139–159.

Gulich D, Zunino L. 2012. The effects of observational correlated noises on multifractal detrended fluctuation analysis. Physica A: Statistical Mechanics and its Applications 391:4100–4110.

Guttorm, T. K., Leppänen, P. H. T., Hämäläinen, J. A., Eklund, K. M., & Lyytinen, H. J. (2011). Newborn event-related potentials predict poorer pre-reading skills in children at risk for dyslexia. Journal of Learning Disabilities, 43(5), 391–401. doi: 10.1177/0022219409345005

Hagiwara, K., Martin, A. D., Nomura, D., & Teubner, T. (2007). Improved predictions for g−2g−2 of the muon and α^{QED}(MZ2). Physics Letters B, 649(2-3), 173–179. doi:10.1016/j.physletb.2007.04.012

Haken, H., Kelso, J., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. Biological Cybernetics, 51(5), 347–356. doi:10.1007/BF00336922

Hänggi P. 2002. Stochastic resonance in biology: How noise can enhance detection of weak signals and help improve biological information processing. ChemPhysChem 3:285–290.

Haselager, P., de Groot, A., & van Rappard, H. (2003). Representationalism vs. anti-representationalism: A debate for the sake of appearance. Philosophical Psychology, 16(1), 5–24. doi:10.1080/0951508032000067761

Hasselman F. 2014a. Predicting Reading Performance from Pre-Literate Speech Perception in Children At-Risk for Dyslexia: Context Relativity of Ordering Relationships and Prospective Prediction. In: Beyond the Boundary. An analysis of verisimilitude and causal ontology of scientific claims: Aetiologies of developmental dyslexia as a case in point.

Hasselman, F. (2013). When the blind curve is finite: dimension estimation and model inference based on empirical waveforms. Frontiers in Physiology, 4(April), 75. doi:10.3389/fphys.2013.00075

Hasselman F. 2014b. Beyond the static phoneme boundary: The nonlinear dynamics of emerging literacy. In: Beyond the Boundary. An analysis of verisimilitude and causal ontology of scientific claims: Aetiologies of developmental dyslexia as a case in point. Aetiologies of developmental dyslexia as a case in point., chapter 5. figshare,.

Hasselman F. The Self-Tuning, Self-Affine Resonator: Radical Embodied Computation as Mental Gymnastics for a Lazy Brain. Frontiers in Human Neuroscience.

Hasselman, Fred and Seevinck, Michael P and Cox, Ralf F. A., Caught in the Undertow: There is Structure Beneath the Ontic Stream (December 21, 2010). Available at SSRN: http://ssrn.com/abstract=2553223

Hay, J., & Drager, K. (2007). Sociophonetics. Annual Review of Anthropology, 36(1), 89–103. doi:10.1146/annurev.anthro.34.081804.120633

Heisenberg, W. (1963). Oral history interview of Werner Heisenberg by Thomas Kuhn. In Archive for the History of Quantum Physics, deposit at Harvard University, Cambridge, MA.

Heisenberg, W. (1971). Physics and beyond; encounters and conversations. New York: Harper & Row.

Hempel, C. (1968). Maximal Specificity and Lawlikeness in Probabilistic Explanation. Philosophy of Science, 35(2), 116–133. Retrieved from http://www.jstor.org/stable/10.2307/186482

Hempel, C., & Oppenheim, P. (1948). Studies in the Logic of Explanation. Philosophy of Science, 15(2), 135–175. doi: 10.2307/185169

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. Neuron, 76(3), 486–502. doi: 10.1016/j.neuron.2012.10.011

Hermens, R. (2011). The problem of contextuality and the impossibility of experimental metaphysics thereof. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 42(4), 214–225. doi: 10.1016/j.shpsb.2011.06.001

Higgins, J., Green, S., & Collaboration, C. (2008). Cochrane Handbook for Systematic Review of Interventions. doi: 10.1002/9780470712184.fmatter

Hogben, L. (1956). The Present Crisis in Statistical Theory. The Incorporated Statistician, 7(1), 3–21. doi:10.2307/2986496

Holden JG, Greijn LT, van Rooij MMJW, Wijnants ML, Bosman AMT. 2014. Dyslexic and skilled reading dynamics are self-similar. Annals of Dyslexia 64:202–221.

Holden JG. (2013). Cognitive Effects as Distribution Rescaling. Ecological Psychology 25:256–266.

Holden, J. G., Choi, I., Amazeen, P. G., & Van Orden, G. C. (2011). Fractal $1/f$ dynamics suggest entanglement of measurement and human performance. Journal of Experimental Psychology. Human Perception and Performance, 37(3), 935–48. doi:10.1037/a0020991

Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. Psychological Review, 116(2), 318–42. doi:10.1037/a0014849

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. Behavioral and Brain Sciences, 24(05), 849–878. doi:doi:10.1017/S0140525X01000103

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. PNAS, 79, 2554–2558.

Hopfield, J. J., & Brody, C. D. (2001). What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. PNAS, 98(3), 1282–7. doi:10.1073/pnas.031567098

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. Biometrical Journal, 50(3), 346–63. doi:10.1002/bimj.200810425

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. Science, 279(5354), 1213–6.

Hsu, H.-C., Fogel, A., & Cooper, R. B. (2000). Infant vocal development during the first 6 months: speech quality and melodic complexity. Infant and Child Development, 9(1), 1–16.

Huss, M., Verney, J., Fosker, T., Mead, N., & Goswami, U. (2010). Music, Rhythm, Rise Time Perception and Developmental Dyslexia: Perception of Musical Meter predicts Reading and Phonology. Cortex, 1–16. doi:10.1016/j.cortex.2010.07.010

Ihlen E A F, Vereijken B. 2013. Multifractal formalisms of human behavior. Human Movement Science 32:633–651.

Ihlen E A F. 2012. Introduction to multifractal detrended fluctuation analysis in Matlab. Frontiers in Physiology 3 JUN:1–51.

Ihlen, E. A. F., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: beyond 1/f(alpha) fluctuation. Journal of Experimental Psychology. General, 139(3), 436–63. doi:10.1037/a0019098

Ioannidis, J. P. A. (2005). Why most published research findings are false. PLoS Medicine, 2(8), e124. doi:10.1371/journal.pmed.0020124

Ioannidis, J. P. A. (2012). Why Science Is Not Necessarily Self-Correcting. Perspectives on Psychological Science, 7(6), 645–654. doi:10.1177/1745691612464056

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: do they arise from a common mechanism? Perception & Psychophysics, 62(4), 874–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10883591

James, W. (1890). The Principles of Psychology (Vol. I & Vol. II). Henry Holt and Company.

James, W. (2008). XI. Letter to Henry Holt. The Letters of William James: 2 Volumes Combined. New York, USA: Cosimo Inc.

Jusczyk, P., Rosner, B., & Cutting, J. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. Perception & Psychophysics, 21(1), 50–54.

Kaiser M, Görner M, Hilgetag CC. 2007. Criticality of spreading dynamics in hierarchical cluster networks without inhibition. New Journal of Physics 9:110–110.

Kaiser M, Hilgetag CC. 2010. Optimal hierarchical modular topologies for producing limited sustained activation of neural networks. Frontiers in neuroinformatics 4:8.

Kantelhardt JW, Zschiegner S A., Koscielny-Bunde E, Havlin S, Bunde A, Stanley HE. 2002. Multifractal detrended fluctuation analysis of nonstationary time series. Physica A: Statistical Mechanics and its Applications 316:87–114.

Kantelhardt JW. 2011. Mathematics of Complexity and Dynamical Systems. In: Meyers RA ed. Mathematics of Complexity and Dynamical Systems. New York, NY: Springer New York, 463–487.

Kantz H, Schreiber T. 2005. Nonlinear Time Series Analysis. Cambridge, UK: Cambridge University Press.

Kello CT, Anderson GG, Holden JG, Van Orden GC. 2008. The Pervasiveness of 1/f Scaling in Speech Reflects the Metastable Basis of Cognition. Cognitive science 32:1217–1231.

Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2007). The emergent coordination of cognitive function. Journal of Experimental Psychology. General, 136(4), 551–68. doi:10.1037/0096-3445.136.4.551

Kello, C. T., Brown, G. D. A., Ferrer-i-Cancho, R., Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., & Van Orden, G. C. (2010). Scaling laws in cognitive sciences. Trends in Cognitive Sciences, 14(5), 223–232.

Kelso J. A. S, Schöner G. 1988. Self-organisation of coordinative movement patterns. Human Movement Science 7:27–46.

Kelso J. A. S. (2012). Multistability and metastability: understanding dynamic coordination in the brain. Philosophical Transactions of the Royal Society B: Biological Sciences 367:906–918.

Kelso, J. A. S. (1986). Pattern formation in speech and limb movements involving many degrees of freedom. Experimental Brain Research 15, pp. 105–128.

Kelso, J. A. S. (1995). Dynamic patterns: The self-organization of brain and behavior. The MIT Press.

Kelso, J., & Schöner, G. (1988). Self-organization of coordinative movement patterns. Human Movement Science, 7(1), 27–46.

Kelso, J., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. Journal of Experimental Psychology: Human Perception and Performance, 10(6), 812.

Kerr, N. (1998). HARKing: Hypothesizing after the results are known. Personality and Social Psychology Review, 2(3), 196–217.

Kievit, R. A, Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. Frontiers in Psychology, 4(August), 513. doi:10.3389/fpsyg.2013.00513

Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher & W. C. Salmon (Eds.), Scientific Explanation (pp. 410–505). Minneapolis: University of Minnesota Press.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams ., R. B., Bahník, Š., Bernstein, M. J., … Nosek, B. a. (2014). Investigating Variation in Replicability. Social Psychology, 45(3), 142–152. doi:10.1027/1864-9335/a000178

Kleinschmidt DF, Jaeger TF. 2014. Robust speech perception: Recognize the familiar , generalize to the similar , and adapt to the novel. Psychological Review in press.

Kloos, H., & Van Orden, G. C. (2009). Soft-assembled mechanisms for the unified theory. In J. P. Spencer, M. Thomas, & J. McClelland (Eds.), Toward a unified theory of development: Connectionism and dynamic systems theory re-considered (pp. 253–267). Oxford: Oxford University Press. Retrieved from http://www.uc.edu/ccrl/documents/Kloos_VanOrden_09.pdf

Koster, C., Been, P. H., Krikhaar, E. M., Zwarts, F., Diepstra, H. D., & Leeuwen, T. H. V. (2005). Differences at 17 Months : Productive Language Patterns in Infants at Familial Risk for Dyslexia and Typically Developing Infants. Journal of Speech, Language, and Hearing Research, 48(2), 426.

Kugler, P. N., Shaw, R. E., Vincente, K. J., & Kinsella-Shaw, J. (1990). Inquiry into intentional systems I: Issues in ecological physics. Psychological Research, 52(2), 98–121. doi:10.1007/BF00877518

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. Perception & Psychophysics, 50(2), 93–107.

Kullback, S., & Leibler, R. (1951). On information and sufficiency. The Annals of Mathematical Statistics, 22(1), 79–86. Retrieved from http://www.jstor.org/stable/10.2307/2236703

Kuznetsov, N. a, & Wallot, S. (2011). Effects of accuracy feedback on fractal characteristics of time estimation. Frontiers in Integrative Neuroscience, 5(October), 62. doi:10.3389/fnint.2011.00062

Kwok T, Smith K a. 2005. Optimization via intermittency with a self-organizing neural network. Neural computation 17:2454–2481.

Kwon O, Moon HT. 2002. Coherence resonance in small-world networks of excitable cells. Physics Letters, Section A: General, Atomic and Solid State Physics 298:319–324.

Ladd, G. (1892). Psychology as So-Called "natural Science". The Philosophical review, 1(1),24–53. retrieved from http://www.jstor.org/stable/10.2307/2175528

Lagarde, J., & Kelso, J. A S. (2006). Binding of movement, sound and touch: multimodal coordination dynamics. Experimental Brain Research, 173(4), 673–88. doi:10.1007/s00221-006-0410-1

Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programmes. In I. Lakatos & A. Musgrave (Eds.), Criticism and the growth of knowledge (p. 170–). Cambridge Univ Press. doi:10.1007/978-94-010-1863-0_14

Lakatos, I. (1974). The role of crucial experiments in science. Studies in History and Philosophy of Science Part A, 4(4), 309–325. doi:10.1016/0039-3681(74)90007-7

Lamb, H. [Kanwal RP.] 1962. The Dynamical Theory of Sound. Dover Pubns.

Landerl, K., & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. Applied Psycholinguistics, 21(2), 243–262. doi:10.1017/S0142716400002058

Lane, H. (1965). The motor theory of speech perception: A critical review. Psychological Review, 72(4), 275

Lasky, R., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT Discrimination by four to six and a half month old infants from spanish environments. Journal of Experimental Child Psychology, (20), 215–225.

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. Review of General Psychology, 15(4), 371–379. doi:10.1037/a0025172

Lebel, E. P., Borsboom, D., Giner-Sorolla, R., Hasselman, F., Peters, K. R., Ratliff, K. a., & Smith, C. T. (2013). PsychDisclosure.org: Grassroots Support for Reforming Reporting Standards in Psychology. Perspectives on Psychological Science, 8(4), 424–432. doi:10.1177/1745691613491437

Li, J. H., Michel, A. N., & Porod, W. (1989). Analysis and synthesis of a class of neural networks: linear systems operating on a closed hypercube. IEEE Transactions on Circuits and Systems, 36(11), 1405–1422.

Liberman A M, Cooper FS, Shankweiler DP, Studdert-Kennedy M. 1967. Perception of the speech code. Psychological review 74:431–461.

Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants. The American Journal of Psychology, 65(4), 497–516.

Lichtwarck-Aschoff, A., Hasselman, F., Cox, R., Pepler, D., & Granic, I. (2012). A characteristic destabilization profile in parent-child interactions associated with treatment efficacy for aggressive children. Nonlinear Dynamics, Psychology, and Life Sciences, 16(3), 353–79.

Lightstone, A., Lightstone, T., & Wilkins, A. (1999). Both coloured overlays and coloured lenses can improve reading fluency, but their optimal chromaticities differ. Ophthalmic and Physiological Optics, 19(4), 279–285. doi:10.1046/j.1475-1313.1999.00442.x

Linell, P. (2005). The written language bias in linguistics: Its nature, origins and transformations (Vol. 5). Psychology Press.

Lisker, L., & Abramson, A. (1971). Distinctive features and laryngeal control. Language, 47(4), 767–785. Retrieved from http://www.jstor.org/stable/10.2307/412155

Little M A, McSharry PE, Roberts SJ, Costello D A E, Moroz IM. 2007. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomedical engineering online 6:23.

Livingstone MS, Rosen GD, Drislane FW, Galaburda a M. 1991. Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. Proceedings of the National Academy of Sciences of the United States of America 88:7943–7947.

Lopes M a, Goltsev a V, Mendes JFF, Resonance SS. 2013. Stochastic resonance as an emergent property of neural networks(仿脑电脑). In: PHYSICS, COMPUTATION, AND THE MIND-ADVANCES AND CHALLENGES AT INTERFACES: Proceedings of the 12th Granada Seminar on Computational and Statistical Physics. 1–6.

Lowie W, Plat R, de Bot K. 2014. Pink Noise in Language Production: A Nonlinear Approach to the Multilingual Lexicon. Ecological Psychology 26:216–228.

Lutz, S. (2012). On a Straw Man in the Philosophy of Science: A Defense of the Received View. HOPOS, 2(1), 77-120.

Lykken, D. T. (1968). Statistical significance in psychological research. Psychological Bulletin, 70(3), 151–9.

Lyytinen, H., Guttorm, T. K., Huttunen, T., Hämäläinen, J., Leppänen, P. H. T., & Vesterinen, M. (2005). Psychophysiology of developmental dyslexia: a review of findings including studies of children at risk for dyslexia. Journal of Neurolinguistics, 18(2), 167–195. doi:10.1016/j.jneuroling.2004.11.001

Mackie, J. (1965). Causes and conditions. American Philosophical Quarterly, 2(4), 245–264. doi:10.2307/20009173

Madurski, C. and LeBel, E. P. (2014). Making sense of the noise: Replication difficulties of Correll's (2008) modulation of 1/f noise in a racial bias task. Psychonomic Bulletin & Review, (2008).

Manis, F. R., & Keating, P. (2005). Speech Perception in Dyslexic Children With and Without Language Impairments. In H. W. Catts & A. G. Kamh (Eds.), The connections between language and reading disabilities (pp. 77–99). New Jersey, US: Lawrence Erlbaum Associates Publishers.

Marinelli, C. V., Angelelli, P., Di Filippo, G., & Zoccolotti, P. (2011). Is developmental dyslexia modality specific? A visual auditory comparison of Italian dyslexics. Neuropsychologia, 363, 1451–1461. doi:10.1016/j.neuropsychologia.2011.02.050

Marmelat V, Delignières D. (2012). Strong anticipation: Complexity matching in interpersonal coordination. Experimental Brain Research 222:137–148.

Marwan N, Carmen Romano M, Thiel M, Kurths J. (2007). Recurrence plots for the analysis of complex systems. Physics Reports 438:237–329.

Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. Speech Communication, 2(1), 15–35.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in Speech and NonSpeech Modes. Cognitive Psychology, 2(2), 131–157.

Mayo, D. G., & Spanos, A. (2006). Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. The British Journal for the Philosophy Science, 57(2), 323–357. doi:10.1093/bjps/axl003

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. Cognition, 95(2), B15–26. doi:10.1016/j.cognition.2004.07.005

Meehl, P. E. (1967). Theory testing in psychology and physics: a methodological paradox. Philosophy of Science, 34, 103–115. doi:10.2307/186099

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46(September 1976), 806–834.

Meehl, P. E. (1990a). Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles That Warrant It. Psychological Inquiry, 1(2), 108–141. doi:10.1207/s15327965pli0102_1

Meehl, P. E. (1990b). Corroboration and Verisimilitude: Against Lakatos's "Sheer Leap of Faith" (No. MCPS–90–01) (Vol. 1990). Minneapolis. Retrieved from http://www.tc.umn.edu/~pemeehl/146CorroborationVerisimilitude.pdf

Meehl, P. E. (1990c). Why Summaries of Research on Psychological Theories Are Often Uninterpretable. Psychological Reports, 66(1), 195-244. doi:10.2466/PR0.66.1.195-244

Meehl, P. E. (1992). Cliometric metatheory: The actuarial approach to empirical, history-based philosophy of science. Psychological Reports, 71(Monograph Supplement 1-V71), 339–467.

Meehl, P. E. (1993). Philosophy of Science: Help or Hindrance? Psychological Reports, 72, 707–733. doi:10.2466/pr0.1993.72.3.707

Meehl, P. E. (1997). The Problem Is Epistemology, Not Statistics: Replace Significance Tests by Confidence Intervals and Quantify Accuracy of Risky Numerical Predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), What if there were no significance tests? (pp. 393–425). New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Meehl, P. E. (2002). Cliometric metatheory: II. Criteria scientists use in theory appraisal and why it is rational to do so. Psychological Reports, 91(2), 339–404.

Meehl, P. E. (2004a). Cliometric metatheory III: Peircean consensus, verisimilitude and asymptotic method. The British Journal for the Philosophy of Science, 55, 615–643.

Meehl, P. E. (2004b). What's in a taxon? Journal of Abnormal Psychology, 113(1), 39–43. doi:10.1037/0021-843X.113.1.39

Merleau-Ponty M. (1962). Phenomenology of perception, trans. C. Smith. London: Routledge.

Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P. 1996. Temporal processing deficits of language-learning impaired children ameliorated by training.

Michell, J. (2009). The psychometricians' fallacy: too clever by half? The British Journal of Mathematical and Statistical Psychology, 62, 41–55. doi:10.1348/000711007X243582

Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: auditory processing or phonological coding? Journal of Experimental Child Psychology, 64(2), 199–231. doi:10.1006/jecp.1996.2343

Moin P, Kim J. 1997. Tackling Turbulence with Supercomputers. Scientific American 276:62–68.

Molenaar, P. C. M. (2008). On the Implications of the Classical ergodic Theorems : Analysis of developmental Processes has to Focus on Intra-Individual Variation. developmental Psychobiology, 50(1), 60–69. doi:10.1002/dev

Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. Current Directions in Psychological Science, 18(2), 112–117. doi:10.1111/j.1467-8721.2009.01619.x

Molfese, D. L. (2000). Predicting dyslexia at 8 years of age using neonatal brain responses. Brain and Language, 72(3), 238–45. doi:10.1006/brln.2000.2287

Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN Is Not a Measure of Orthographic Processing. Evidence From the Asymmetric German Orthography. Scientific Studies of Reading, 13(1), 1–25. doi:10.1080/10888430802631684

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2012). Language experienced in utero affects vowel perception after birth: a two-country study. Acta paediatrica, 156–160. doi:10.1111/apa.12098

Moores, E. (2004). Deficits in dyslexia: barking up the wrong tree? Dyslexia, 10(4), 289–98. doi:10.1002/dys.277

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? Cognition, 7(4), 323–331.

Munhall, K. G., Löfqvist, A., & Kelso, J. (1994). Lip-larynx coordination in speech: effects of mechanical perturbations to the lower lip. The Journal of the Acoustical Society of America, 95(6), 3605–16.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. Journal of Mathematical Psychology, 47(1), 90–100. doi:10.1016/S0022-2496(02)00028-7

Nagarajan, S. S., Wang, X., Merzenich, M. M., Schreiner, C. E., Johnston, P., Jenkins, W. M., Miller, S., et al. (1998). Speech modifications algorithms used for training language learning-impaired children. IEEE transactions on rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society, 6(3), 257–68. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9749903

Nam, H., Goldstein, L., & Saltzman, E. L. (2009). Self-organization of Syllable Structure: a Coupled oscillator model. In F. Pellegrino, E. Marisco, & I. Chitoran (Eds.), Approaches to phonological complexity (pp. 1–29). Berlin/New York: Mouton de Gruyter. Retrieved from http://sail.usc.edu/~lgoldste/ArtPhon/Papers/Week 11/Nam_Gold_Saltz.pdf

Nazzi, T. (2003). Perception and acquisition of linguistic rhythm by infants. Speech Communication, 41(1), 233–243. doi: 10.1016/S0167-6393(02)00106-1

Neuhoff, N., Bruder, J., Bartling, J., Warnke, A., Remschmidt, H., Müller-Myhsok, B., & Schulte-Körne, G. (2012). Evidence for the late MMN as a neurophysiological endophenotype for dyslexia. PloS one, 7(5), e34909. doi:10.1371/journal.pone.0034909

Neyman, J. (1969). Statistical Problems in Science. The Symmetric Test of a Composite Hypothesis. Journal of the American Statistical Association, 64(328), 1154–1171. doi: 10.1080/01621459.1969.10501047

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. Psychological Methods, 5(2), 241–301.

Nicolson, R. I., & Fawcett, A. J. (2006). Do cerebellar deficits underlie phonological problems in dyslexia? Developmental Science, 9(3), 259–62; discussion 265–9. doi:10.1111/j.1467-7687.2006.00486.x

Nicolson, R. I., Fawcett, A. J., & Dean, P. (2001). A TINS debate - Hindbrain versus the forebrain: a case for cerebellar deficit in developmental dyslexia. TRENDS in Neurosciences, 24(9), 508–511.

Nicolson, R. I., Fawcett, A. J., Brookes, R. L., & Needle, J. (2010). Procedural learning and dyslexia. Dyslexia, 16(3), 194–212. doi:10.1002/dys

Nieber, R., & Wollock, J. (1997). The Historical Meaning of the Crisis in Psychology: A Methodological Investigation. In r. rieber & J. Wollock (eds.), The Collected Works of L. S. Vygotsky Se - 17. Springer uS. doi: 10.1007/978-1-4615-5893-4_17

Nittrouer, S., Munhall, K. G., Kelso, J., Tuller, B., & Harris, K. S. (1988). Patterns of interarticulator phasing and their relation to linguistic structure. The Journal of the Acoustical Society of America, 84(5), 1653–1661. Retrieved from http://link.aip.org/link/?JASMAN/84/1653/1

Nix, T. W., & Barnette, J. J. (1998). The Data Analysis Dilemma: Ban or Abandon . A Review of Null Hypothesis Significance Testing. Research in the Schools, 5(2), 3–14.

Noordenbos, M. W., Segers, E., Serniclaes, W., Mitterer, H., & Verhoeven, L. (2012a). Allophonic mode of speech perception in Dutch children at risk for dyslexia: A longitudinal study. Research in Developmental Disabilities, 33(5), 1469–1483. doi:10.1016/j.ridd.2012.03.021

Noordenbos, M. W., Segers, E., Serniclaes, W., Mitterer, H., & Verhoeven, L. (2012b). Neural evidence of allophonic perception in children at risk for dyslexia. Neuropsychologia, 50(8), 2010–7. doi:10.1016/j.neuropsychologia.2012.04.026

Nottale, L. (2010). Scale Relativity and Fractal Space-Time: Theory and Applications. Foundations of Science, 15(2), 101–152. doi:10.1007/s10699-010-9170-2

Novalis. (2008). Dialogen und Monolog. In H. Balmes (ed.), F. Hasselman [Quoted paragraph] (Trans.), Gesammelte Werke (pp. 439–451). Frankfurt am Main, Germany: Fischer Taschenbuch Verlag.

Nowak, A., & Vallacher, R. R. (2003). Synchronization Dynamics in Close Relationships: Coupled Logistic Maps as a Model of Interpersonal Phenomena. In W. Klonowski (Ed.), Frontiers on nonlinear dynamics: Vol. 2. From quanta to societies (pp. 165–180). Berlin: Pabst Science Publishers.

Nunnally, J. (1960). The Place of Statistics in Psychology. Educational and Psychological Measurement, 20(4), 641–650. doi: 10.1177/001316446002000401

O'Brien, B. A, Wolf, M., & Lovett, M. W. (2012). A taxometric investigation of developmental dyslexia subtypes. Dyslexia, 18(1), 16–39. doi:10.1002/dys.1431

Ono H, Weber EU. 1981. Journal of experimental psychology: human perception and performance. Journal of experimental psychology. Human perception and performance 7:937–947.

Onofrei M, Hunt J, Siemienczuk J, Touchette DR, Middleton B. 2004. A first step towards translating evidence into practice: Heart failure in a community practice-based research network. Informatics in Primary Care 12:139–145.

Overy, K. (2003). Dyslexia and music. From timing deficits to musical intervention. Annals of the New York Academy of Sciences, 999, 497–505.

Ozer M, Uzuntarla M, Kayikcioglu T, Graham LJ. 2008. Collective temporal coherence for subthreshold signal encoding on a stochastic small-world Hodgkin-Huxley neuronal network. Physics Letters, Section A: General, Atomic and Solid State Physics 372:6498–6503.

Partanen, E., Kujala, T., Naatanen, R., Liitola, a., Sambeth, a., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. Proceedings of the National Academy of Sciences, 110(37). doi:10.1073/pnas.1302159110

Pasquini ES, Corriveau KH, Goswami U. 2007. Auditory Processing of Amplitude Envelope Rise Time in Adults Diagnosed With Developmental Dyslexia. Scientific Studies of Reading 11:259–286.

Perkell JS, Guenther FH, Lane H, Matthies ML, Stockmann E, Tiede M, Zandipour M. (2004a). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. The Journal of the Acoustical Society of America 116:2338–2344.

Perkell JS, Matthies ML, Tiede M, Lane H, Zandipour M, Marrone N, Stockmann E, Guenther FH. (2004b). The distinctness of speakers' /s/-/S/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. Journal of speech, language, and hearing research: JSLHR 47:1259–1269.

Petersen, K. (1996). Ergodic theorems and the basis of science. Synthese, 108(2), 171–183. doi: 10.1007/BF00413496

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: movement influences infant rhythm perception. Science (New York, N.Y.), 308(5727), 1430. doi:10.1126/science.1110922

Pihko, E., Leppänen, P. H. T., Eklund, K. M., Cheour, M., Guttorm, T. K., & Lyytinen, H. J. (1999). Cortical responses of infants with and without a genetic risk for dyslexia: I. Age effects. Neuroreport, 10(5), 901.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. D. C. (2012). nlme: Linear and Nonlinear Mixed Effects Models. Retrieved from http://www.r-project.org/

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. The Journal of the Acoustical Society of America, 55(2), 328–33.

Platt, J. (1964). Strong Inference. Science, 146(3642), 347–353.

Poincaré, H. (1905). Science and Hypothesis. new york: The Walter Scott Publishing Co., LTd.

Popper, K. (1934\1959). The Logic of Scientific Discovery. (J. Freed & L. Freed, Trans.). Hutchinson & Co.

Port, R. F. (2007). How are words stored in memory? Beyond phones and phonemes. New Ideas in Psychology, 25(2), 143–170. doi:10.1016/j.newideapsych.2007.02.001

Port, R. F. (2008a). All is prosody: Phones and phonemes are the ghosts of letters. Proc 4th Internal Conf Speech Prosody (pp. 7–16). Retrieved from http://www.isca-speech.org/archive/sp2008/papers/sp08_007.pdf

Port, R. F. (2008b). No role for phonology in speech processing: Taking a strong position. Proceedings of Meetings on Acoustics (Vol. 4, p. 060008). Retrieved from http://www.cs.indiana.edu/~port/pap/Port.no.role.for.phonology.2009.POMA.pdf

Port, R. F. (2010). Language as a Social Institution: Why Phonemes and Words Do Not Live in the Brain. Ecological Psychology, 22(4), 304–326. doi: 10.1080/10407413.2010.517122

Port, R. F. (2010b). The reality of phonological forms: a rejoinder. Language Sciences, 32(1), 60–62. doi:10.1016/j.langsci.2009.10.016

Port, R. F., & Leary, A. P. (2005). against formal phonology. Language, 81(4), 927–964.

Porter RJ, Hogue DM. 1998. Nonlinear dynamical systems in speech perception and production. Nonlinear Dynamics, Psychology, and Life Sciences 2:95–131.

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. NeuroImage, 19(3), 473–481. doi:10.1016/S1053-8119(03)00084-3

Proverbio, A. M., Manfredi, M., Zani, A., & Adorni, R. (2013). Musical expertise affects neural bases of letter recognition. Neuropsychologia, 51(3), 538–49. doi:10.1016/j.neuropsychologia.2012.12.001

Pugh, K. R., Mencl, W. E., Jenner, a R., Katz, L., Frost, S. J., Lee, J. R., … Shaywitz, B. A. (2001). Neurobiological studies of reading and reading disability. Journal of Communication Disorders, 34(6), 479–92.

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H. T., Poikkeus, A.-M., Tolvanen, A., et al. (2007). Very early phonological and language skills: estimating individual risk of reading disability. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 48(9), 923–31. doi:10.1111/j.1469-7610.2007.01763.x

R Core Team. (2012). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org/

Rabash J, Charlton C, Brown WJ, Healy M, Cameron B. 2009. MLwiN (Version 2.2) [Computer Program].

Ramus F. 2004. Neurobiology of dyslexia: A reinterpretation of the data. Trends in Neurosciences 27:720–726.

Ramus, F. (2003). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? Current Opinion in Neurobiology, 13(2), 212–218. doi:10.1016/S0959-4388(03)00035-7

Ramus, F. (2003b). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. Brain, 126(4), 841–865. doi:10.1093/brain/awg076

Ramus, F. (2004). Neurobiology of dyslexia: A reinterpretation of the data. TRENDS in Neurosciences, 27(12), 720–726. doi:10.1016/j.tins.2004.10.004

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? The Quarterly Journal of Experimental Psychology, 61(1), 129–141. doi:10.1080/17470210701508822

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. Science, 288(5464), 349–51.

Ramus, F., Nespor, M., & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. Cognition, 75(1), AD3–AD30.

Ramus, F., Pidgeon, E., & Frith, U. (2003). The relationship between motor control and phonology in dyslexic children. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 44(5), 712–22.

Ramus, F., White, S., & Frith, U. (2006). Weighing the evidence between competing theories of dyslexia. Developmental Science, 9(3), 265–269. doi:10.1111/j.1467-7687.2006.00488.x

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2009). MLwiN Version 2.1. Bristol: Centre for Multilevel Modelling, University of Bristol.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. Psychological science in the public interest, 2(2), 31.

Read, C., Zhang, Y. F., Nie, H. Y., & Ding, B. Q. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. Cognition, 24(1-2), 31–44.

Remak, R. (1855). untersuchungen über die entwickelung der Wirbelthiere. Berlin: G. reimer.

Reynolds, D., & Nicolson, R. I. (2007). Follow-up of an exercise-based treatment for children with reading difficulties. Dyslexia, 13(2), 78–96. doi:10.1002/dys

Reynolds, D., Nicolson, R. I., & Hambly, H. (2003). Evaluation of an exercise-based treatment for children with reading difficulties. Dyslexia, 9(1), 48–71; discussion 46–7. doi:10.1002/dys.235

Richardson DC, Dale R, Kirkham NZ. 2007. The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue: Research article. Psychological Science 18:407–413.

Richardson, U., Leppänen, P. H. T., Leiwo, M., & Lyytinen, H. J. (2003). Speech perception of infants with high familial risk for dyslexia differ at the age of 6 months. Developmental Neuropsychology, 23(3), 385–97.

Rigoli LM, Holman D, Spivey MJ, Kello CT. 2014. Spectral convergence in tapping and physiological fluctuations: coupling and independence of 1/f noise in the central and autonomic nervous systems. Frontiers in Human Neuroscience 8:1–10.

Riley M a, Van Orden GC. 2005. Tutorials in contemporary nonlinear methods for the behavioral sciences. National Science Foundation.

Ring, K. (1967). Experimental social psychology: Some sober questions about some frivolous values. Journal of Experimental Social Psychology, 3(2), 113–123. doi:10.1016/0022-1031(67)90016-9

Rochelle, K. S. H., & Talcott, J. B. (2006). Impaired balance in developmental dyslexia? A meta-analysis of the contending evidence. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 47(11), 1159–66. doi:10.1111/j.1469-7610.2006.01641.x

Rome, E. (2008). MACS - Final Activity Report. Methodology. Retrieved from http://www.macs-eu.org/deliverables/MACS_FAR_071130v3(19c)-MR.pdf

Rosnow, R., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44(10), 1276–1284.

Rozeboom, W. (1960). The fallacy of the null-hypothesis significance test. Psychological Bulletin, 57, 416–28.

Rozeboom, W. (1982). Let's Dump Hypothetico-Deductivism for the Right Reasons. Philosophy of Science, 49(4), 637–647. doi:10.2307/187171

Salmon, W. C. (1978). Why ask, "Why?"? An inquiry concerning scientific exploration. In Fifty-second Annual Pacific Meeting of the American Philosophical Association (pp. 683–705). San Francisco, CA.

Salmon, W. C. (1984). Scientific Explanation and Causal Structure of the World. Princeton: Princeton University Press.

Salmon, W. C. (1990). Scientific explanation: Causation and Unification. Critica: Revista Hispanoamericana de Filosofia, XXII(66), 3–23. Retrieved from http://www.arkiv.certec.lth.se/kk/dokument/causalityandexplanation.pdf

Salmon, W. C. (1999). Scientific explanation. In M. H. Salmon (Ed.), Introduction to the philosophy of science (pp. 7–41). Indianapolis, IN, USA: Hackett Publishing Company, Inc.

Saltzman, E. L., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. Human Movement Science, 19(4), 499–526. doi:10.1016/S0167-9457(00)00030-0

Saltzman, E. L., & Munhall, K. G. (1989). A Dynamical Approach to Gestural Patterning in Speech Production (No. SR-99 / 100, 38-68) (pp. 38–68).

Sannino, A., & Sutter, B. (2011). Cultural-historical activity theory and interventionist methodology: Classical legacy and contemporary developments. Theory & Psychology, 21(5), 557–570. doi:10.1177/0959354311414969

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. Psychological Methods, 17(4), 551–66. doi:10.1037/a0029487

Schinkel S, Dimigen O, Marwan N. 2008. Selection of recurrence threshold for signal detection. European Physical Journal: Special Topics 164:45–53.

Schinkel S, Marwan N, Dimigen O, Kurths J. 2009. Confidence bounds of recurrence-based complexity measures. Physics Letters, Section A: General, Atomic and Solid State Physics 373:2245–2250.

Schinkel S, Marwan N, Kurths J. 2009. Brain signal analysis based on recurrences. Journal of Physiology Paris 103:315–323.

Schinkel, S., Marwan, N., & Kurths, J. (2007). Order patterns recurrence plots in the analysis of ERP data. Cognitive Neurodynamics, 1(4), 317–25. doi:10.1007/s11571-007-9023-z

Schinkel, S., Marwan, N., & Kurths, J. (2009). Journal of Physiology - Paris Brain signal analysis based on recurrences. Journal of Physiology - Paris, 103(6), 315–323. doi:10.1016/j.jphysparis.2009.05.007

Scholz JP, Schöner G. 1999. The uncontrolled manifold concept: Identifying control variables for a functional task. Experimental Brain Research 126:289–306.

Schöner G, Kelso J A. S. 1988. Dynamic pattern generation in behavioral and neural systems. Science (New York, N.Y.) 239:1513–1520.

Schöner, G., & Dineva, E. (2007). Dynamic instabilities as mechanisms for emergence. Developmental Science, 10(1), 69–74. doi:10.1111/j.1467-7687.2007.00566.x

Schöner, G., & Kelso, J. A. S. (1988). Dynamic Pattern Generation in Behavioral and Neural Systems. Science, 239(4847), 1513–1520.

Schöner, G., Haken, H., & Kelso, J.. A. S. (1986). A Stochastic Theory of Phase Transitions in Human Hand Movement. Biological Cybernetics, 53(4), 247–257.

Schouten, B., Gerrits, E., & Van Hessen, A. J. (2003). The end of categorical perception as we know it. Speech Communication, 41(1), 71–80. doi:10.1016/S0167-6393(02)00094-8

Schouten, M. E. H. (1980). The case against a speech mode of perception. Acta Psychologica, 44(1), 71–98.

Schrödinger, E. (1935). Die gegenwärtige Situation in der Quantenmechanik. Naturwissenschaften, 23, 807–849.Retrieved from http://www.springerlink.com/index/ HL6541202uKn4741.pdf

Scripture EW. (1891). Psychology as philosophy. Mind 63:306–326.

Segers E, Verhoeven L. 2005. Effects of lengthening the speech signal on auditory word discrimination in kindergartners with SLI. Journal of Communication Disorders 38:499–514.

Segers, E., & Verhoeven, L. (2005). Effects of lengthening the speech signal on auditory word discrimination in kindergartners with SLI. Journal of Communication Disorders, 38(6), 499–514. doi:10.1016/j.jcomdis.2005.04.003

Segers, E., Hasselman, F., Verhoeven, L., & De Graaff, S. (2004). Spraakmanipulatie en educatieve software ter stimulering van het fonologisch bewustzijn van kleuters met een risico op leesproblemen. Stem, Spraak en Taalpathologie, 12(3), 187–202. Retrieved from http://hdl.handle.net/2066/64573

Sellars, W. (1956). Empiricism and the Philosophy of Mind. In H. Feigl & H. Scriven (Eds.), Minnesota Studies in the Philosophy of Science: The Foundations of Science and the Concepts of Psychology and Psychoanalysis (Vol. 1, pp. 253–329). Minnesota: University of Minnesota Press.

Serniclaes W. 2006. Categorical perception of speech sounds and dyslexia. Current psychology letters. … 1:1–8.

Serniclaes, W. (2006). Allophonic perception in developmental dyslexia: Origin, reliability and implications of the categorical perception deficit. Written Language and Literacy, 9(1), 135–152.

Serniclaes, W. (2006). Allophonic perception in developmental dyslexia: Origin, reliability and implications of the categorical perception deficit. Written Language and Literacy, 9(1), 135–152. Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=18245907

Serniclaes, W., & Sprenger-Charolles, L. (2006). Categorical perception of speech sounds and dyslexia. Current Psychology Letters, 1(10, Vol. 1, 2003), 1–9. Retrieved from http://cpl.revues.org/index379.html

Serniclaes, W., Sprenger-Charolles, L., Carre, R., & Demonet, J. F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. Journal of Speech, Language, and Hearing Research, 44(2), 384. doi:10.1111/j.1460-9568.2011.07697.x

Serniclaes, W., Van Heghe, S., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. Journal of Experimental Child Psychology, 87(4), 336–61. doi:10.1016/j.jecp.2004.02.001

Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. Cognition, 98(2), B35–44. doi:10.1016/j.cognition.2005.03.002

Shaw, R. E., & Turvey, M. T. (1999). Ecological foundations of cognition: II. Degrees of freedom and conserved quantities in animal-environment systems. Journal of Consciousness Studies, 6(11-12), 111–123.

Shaw, R., & Kinsella-Shaw, J. (1988). Ecological mechanics: A physical geometry for intentional constraints. Human Movement Science, 7, 155–200.

Shaw, R., Flascher, O., & Mace, W. (1996). Dimensions of Event Perception. In Handbook of perception and action (Vol. 1, pp. 345–295). New York: Academic Press.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. New England Journal of Medicine, 326(3), 145–150.

Shockley, K., Baker, A. a, Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. Journal of Experimental Psychology. Human perception and performance, 33(1), 201–8. doi: 10.1037/0096-1523.33.1.201

Silvén, M., Poskiparta, E., Niemi, P., & Voeten, M. (2007). Precursors of reading skill from infancy to first grade in Finnish: Continuity and change in a highly inflected language. Journal of Educational Psychology, 99(3), 516–531. doi: 10.1037/0022-0663.99.3.516

Simons, D. J. (2014). The Value of Direct Replication. Perspectives on Psychological Science, 9(1):76–80.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. Journal of Experimental Psychology. General, 143(2), 534–47. doi:10.1037/a0033242

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science, July, 265–269.

Smit-Glaudé, S. W. D., van Strien, J. W., Licht, R., & Bakker, D. J. (2005). Neuropsychological intervention in kindergarten children with subtyped risks of reading retardation. Annals of Dyslexia, 55(2), 217–45. doi:10.1007/s11881-005-0012-9

Smith ZM, Delgutte B, Oxenham AJ. 2002. Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87–90.

Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. Developmental Review, 25(3-4), 278–298. doi:10.1016/j.dr.2005.11.001

Smith, L. B., & Gasser, M. (2005). The development of embodied cognition: six lessons from babies. Artificial life, 11(1-2), 13–29. doi:10.1162/1064546053278973

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. Trends in Cognitive Sciences, 7(8), 343–348. doi: 10.1016/S1364-6613(03)00156-6

Smotherman, W., & Robinson, S. R. (1988). The uterus as environment. The ecology of fetal behavior. In E. M. Blass (Ed.), Handbook of behavioral neurobiology, Vol. 9 (pp. 149–196). Plenum Publishing Corporation. Retrieved from http://www.psychology.uiowa.edu/Faculty/Robinson/SRR_Reprints/88HBN.pdf

Snowling, M. J. (2001). From language to reading and dyslexia. Dyslexia, 7(1), 37–46. doi:10.1002/dys.185

Spencer, H. (1855). The Principles of Psychology. London: Longman, Brown, Green and Longmans. Retrieved from: http://oll.libertyfund.org/titles/1394

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. Developmental Science, 6(4), 392–412. doi:10.1111/1467-7687.00295

Spivey, M. J., & Anderson, S. E., & Dale, R. (2009). The phase transition in human cognition, New Mathematics and Natural Computation, 5, 197–220.

Spivey, M. J., & Dale, R. (2006). Continuous Dynamics in Real-Time Cognition. Current Directions in Psychological Science, 15(5), 207–211. doi:10.1111/j.1467-8721.2006.00437.x

Stanovich, K. E. (1985). Explaining the variance in reading ability in terms of psychological processes: What have we learned? Annals of Dyslexia, 35(1), 67–96. doi:10.1007/BF02659181

Stanovich, K. E. (1988). The right and wrong places to look for the cognitive locus of reading disability. Annals of Dyslexia, 38(1), 154–177. doi:10.1007/BF02648254

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. Psychological Methods, 9(2), 164–82. doi:10.1037/1082-989X.9.2.164

Stein J, Walsh V. (1997). To see but not to read; the magnocellular theory of dyslexia. Trends in Neurosciences 20:147–152.

Stein, J. F. (2001). The magnocellular theory of developmental dyslexia. Dyslexia, 7(1), 12–36. doi:10.1002/dys.186

Stein, J. F., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. Trends in neurosciences, 20(4), 147–152. doi:10.1016/S0166-2236(96)01005-3

Stephen, D. G., Anastas, J. R., & Dixon, J. a. (2012). Scaling in cognitive performance reflects multiplicative multifractal cascade dynamics. Frontiers in Physiology, 3(April), 1–6. doi:10.3389/fphys.2012.00102

Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: Entropy, Action and Cognition. Journal of Experimental Psychology: Human Perception and Performance, 35(6), 1811–1832.

Stetson RH. 1951. Motor phonetics: A study of speech movements in action. In: Kelso JAS, Munhall KG eds. R. H. Stetson's motor phonetics: A retrospective edition. Boston, MA: College Hill,.

Stoodley, C. J., & Stein, J. F. (2012). Cerebellar Function in Developmental Dyslexia. Cerebellum (London, England). doi: 10.1007/s12311-012-0407-1

Stoodley, C. J., Fawcett, A. J., Nicolson, R. I., & Stein, J. F. (2005). Impaired balancing ability in dyslexic children. Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale, 167(3), 370–80. doi:10.1007/s00221-005-0042-x

Strait, D. L., Parbery-Clark, A., O'Connell, S., & Kraus, N. (2013). Biological impact of preschool music classes on processing speech in noise. Developmental Cognitive Neuroscience, 6C, 51–60. doi:10.1016/j.dcn.2013.06.003

Stroebe, W. and Strack, F. (2014). The Alleged Crisis and the Illusion of Exact Replication. Perspectives on Psychological Science, 9(1):59–71.

Tallal P, Miller S, Fitch RH. 1993. Neurobiological basis of speech: a case for the preeminence of temporal processing. Annals of the New York Academy of Sciences 682:27–47.

Tallal P, Miller SL, Bedi G, Byma G, Wang X, Nagarajan SS, Schreiner C, Jenkins WM, Merzenich MM. 1996. Language comprehension in language-learning impaired children improved with acoustically modified speech.

Tallal P, Piercy M. 1974. Developmental aphasia: rate of auditory processing and selective impairment of consonant perception. Neuropsychologia 12:83–93.

Tallal P. 1976. Rapid auditory processing in normal and disordered language development. Journal of speech and hearing research 19:561–571.

Tallal, P. (2004). Improving language and literacy is a matter of time. Nature Reviews. Neuroscience, 5(9), 721–8. doi: 10.1038/nrn1499

Tallal, P., & Piercy, M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia. Neuropsychologia, 11(4), 389–98.

Temple E, Poldrack R A, Protopapas A, Nagarajan S, Salz T, Tallal P, Merzenich MM, Gabrieli JD. 2000. Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. Proceedings of the National Academy of Sciences of the United States of America 97:13907–13912.

The MathWorks. 2012. Matlab R2012a (version 7.14.0.739) [Computer Software].

Theiler J. 1990. Estimating fractal dimension. Journal of the Optical Society of America A 7:1055.

Thelen, E. (1996). Motor development: A new synthesis. Progress in child Psychiatry and child development.

Thelen, E. (2000). Motor development as foundation and future of developmental psychology. International Journal of Behavioral Development, 24(4), 385–397. doi:10.1080/016502500750037937

Thelen, E., Kelso, J., & Fogel, A. (1987). Self-organizing systems and infant motor development. Developmental Review, 7(1), 39–65.

Thelen, E., Schöner, G., & Scheier, C. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. Behavioral and Brain Sciences, 1–86.

Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. Journal of Physiology, Paris, 102(1-3), 120–9. doi:10.1016/j.jphysparis.2008.03.007

Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: further tests with a non-human primate. Developmental Science, 8(1), 26–35. doi:10.1111/j.1467-7687.2005.00390.x

Titchener, E. (1893). Two recent Criticisms of "Modern" Psychology. The Philosophical review, 2(4), 450–458. retrieved from http://www.jstor.org/stable/10.2307/2175721

Tononi, G., Edelman, G. M., & Sporns, O. (1998). Complexity and coherency: integrating information in the brain. Trends in Cognitive Sciences, 2(12), 474–84.

Torre, K., & Wagenmakers, E.-J. (2009). Theories and models for 1/f(beta) noise in human movement science. Human Movement Science, 28(3), 297–318. doi:10.1016/j.humov.2009.01.001

Torres J-L, Ruiz MG. 1996. Stochastic resonance and the homoeopathic effect. British Homoeopathic journal 85:134–140.

Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. Psychological Review, 110(3), 526–535. doi:10.1037/0033-295X.110.3.526

Tripepi, G., Jager, K. J., Dekker, F. W., & Zoccali, C. (2009). Diagnostic methods 2: receiver operating characteristic (ROC) curves. Kidney International, 76(3), 252–6. doi:10.1038/ki.2009.171

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. Child Development, 75(4), 1067–84. doi:10.1111/j.1467-8624.2004.00726.x

Tukey, J. (1960a). Conclusions vs Decisions. Technometrics, 2(4), 423–433.

Tukey, J. (1960b). Where Do We Go From Here? Journal of the American Statistical Association, 55(289), 80–93.

Tuller, B. (2004). Categorization and learning in Speech Perception as Dynamical Processes. Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences.

Tuller, B., & Kelso, J. (1991). The production and perception of syllable structure. Journal of Speech and Hearing Research, 34(3), 501–8.

Tuller, B., Case, P., Ding, M., & Kelso, J. (1994). The nonlinear dynamics of speech categorization. Journal of Experimental Psychology: Human Perception and Performance, 20(1), 3–16.

Turvey, M. T. (2007). Action and perception at the level of synergies. Human Movement Science, 26(4), 657–97. doi: 10.1016/j.humov.2007.04.002

Turvey, M. T., & Carello, C. (2012). On Intelligence From First Principles: Guidelines for Inquiry Into the Hypothesis of Physical Intelligence (PI). Ecological Psychology, 24(1), 3–32. doi:10.1080/10407413.2012.645757

Uzuntarla M. 2013. Inverse stochastic resonance induced by synaptic background activity with unreliable synapses. Physics Letters A 377:2585–2589.

Vaessen, A., Gerretsen, P., & Blomert, L. (2009). Naming problems do not reflect a second independent core deficit in dyslexia: double deficits explored. Journal of Experimental Child Psychology, 103(2), 202–21. doi:10.1016/j.jecp.2008.12.004

Van Alphen, P., De Bree, E., Gerrits, E., De Jong, J., Wilsenach, C., & Wijnen, F. (2004). Early language development in children with a genetic risk of dyslexia. Dyslexia, 10(4), 265–88. doi:10.1002/dys.272

Van Beinum, F. J., Schwippert, C. E., Been, P. H., Van Leeuwen, T. H., & Kuijpers, C. T. L. (2005). Development and application of a /bAk/–/dAk/ continuum for testing auditory perception within the Dutch longitudinal dyslexia study. Speech Communication, 47(1-2), 124–142. doi:10.1016/j.specom.2005.04.003

Van den Bos KP, Lutje Spelberg HC, Scheepstra a JM, de Vries JR. 1994. De Klepel - Verantwoording, Diagnostiek en Behandeling [Nonword Reading Test]. Nijmegen, The Netherlands: Berkhout.

Van den Bos, K. P., lutje Spelberg, H. C., & Eleveld, M. (2004). Voorspelling van leessnelheid van leerlingen in groep 3 met tests afgenomen in groep 2 en 3 [Prediction of reading speed of children in grade 1 with tests administered in the second kindergarten year and in grade 1]. Tijdschrift voor Orthopedagogiek, 43, 312–324.

Van Herten, M., Pasman, J., Van Leeuwen, T. H., Been, P. H., Van der Leij, A., Zwarts, F., & Maassen, B. (2008). Differences in AERP responses and atypical hemispheric specialization in 17-month-old children at risk of dyslexia. Brain Research, 1201, 100–5. doi:10.1016/j.brainres.2008.01.060

Van Leeuwen, T., Been, P., Kuijpers, C., Zwarts, F., Maassen, B., & Van der Leij, A. (2006). Mismatch response is absent in 2-month-old infants at risk for dyslexia. Neuroreport, 17(4), 351–5. doi:10.1097/01.wnr.0000203624.02082.2d

Van Leeuwen, T., Been, P., Kuijpers, C., Zwarts, F., Maassen, B., van der Leij, A., & Leeuwen, T. Van. (2006). Mismatch response is absent in 2-month-old infants at risk for dyslexia. Neuroreport, 17(4), 351–5. doi:10.1097/01.wnr.0000203624.02082.2d

Van Leeuwen, T., Been, P., Van Herten, M., Zwarts, F., Maassen, B., & Van der Leij, A. (2008). Two-month-old infants at risk for dyslexia do not discriminate /bAk/ from /dAk/: A brain-mapping study. Journal of Neurolinguistics, 21(4), 333–348. doi: 10.1016/j.jneuroling.2007.07.004

Van Lieshout, P. H. H. M. (2004). Dynamical systems theory and its application in speech. In B. Maassen, R. Kent, H. F. M. Peters, P. Peters, P. Van Lieshout, & W. Hulstijn (Eds.), Speech motor control in normal and disordered speech (Vol. 03, pp. 51–82). Oxford: Oxford University Press.

Van Orden, G. C. (1987). A ROWS is a ROSE: spelling, sound, and reading. Memory & cognition, 15(3), 181–98.

Van Orden, G. C., & Kloos, H. (2003). The module mistake. Cortex, 39(1), 164–166.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2005). Human cognition and 1/f scaling. Journal of Experimental Psychology. General, 134(1), 117–23. doi:10.1037/0096-3445.134.1.117

Van Orden, G. C., Holden, J. G., & Turvey, M. T.,(2003). Self-organization of cognitive performance. Journal of Experimental Psychology: General General, 132(3), 331–350. doi:10.1037/0096-3445.132.3.331

Van Orden, G. C., Hollis, G., & Wallot, S. (2012). The blue-collar brain. Frontiers in Physiology, 3(June), 1–12. doi:10.3389/fphys.2012.00207

Van Orden, G. C., Kloos, H., & Wallot, S. (2009). Living in the pink: Intentionality, wellbeing, and complexity. Handbook of the Philosophy of Science. Volume 10: Philosophy of Complex Systems (Vol. 10). Amsterdam: Elsevier B.V

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? Cognitive Science, 25(1), 111–172. doi:10.1207/s15516709cog2501_5

van Rooij, M. M. J. W., Nash, B. a, Rajaraman, S., & Holden, J. G. (2013). A fractal approach to dynamic inference and distribution analysis. Frontiers in Physiology, 4(January), 1. doi:10.3389/fphys.2013.00001

Van Rullen, R., & Koch, C. (2003). Is perception discrete or continuous? Trends in Cognitive Sciences, 7(5), 207–213. doi: 10.1016/S1364-6613(03)00095-0

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? Journal of child psychology and psychiatry, and allied disciplines, 45(1), 2–40.

Verhoeven L. 1995. Drie-minuten-toets. Handleiding [three-minutes-test. Manual]. Arnhem, The Netherlands: CITO.

Virchow, R (1855). Cellular-Pathologie. Archiv für Pathologische Anatomie und Physiologie und für Klinische Medicin, 8(1), 3–39. doi:10.1007/BF01935312

Von Hofsten, C. (2004). An action perspective on motor development. Trends in cognitive sciences, 8(6), 266–72. doi: 10.1016/j.tics.2004.04.002

Von neumann, J. (1932). Mathematical Foundations of Quantum Mechanics (1996 ed.). Princeton, new Jersey: Princeton university Press.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of 1/falpha noise in human cognition. Psychonomic Bulletin & Review, 11(4), 579–615.

Wagenmakers, E.-J., Molenaar, P. C. M., Grasman, R., Hartelman, P., & Van der Maas, H. L. J. (2005). Transformation invariant stochastic catastrophe theory. Physica D: Nonlinear Phenomena, 211(3-4), 263–276. doi:10.1016/j.physd.2005.08.014

Wagenmakers, E.-J., van der Maas, H. L. J., & Farrell, S. (2012). Abstract concepts require concrete models: why cognitive scientists have not yet embraced nonlinearly coupled, dynamical, self-organized critical, synergistic, scale-free, exquisitely context-sensitive, interaction-dominant, multifractal, interdependent b. Topics in Cognitive Science, 4(1), 87–93; discussion 94–102. doi:10.1111/j.1756-8765.2011.01164.x

Wallot S, Hollis G, van Rooij M. 2013. Connected Text Reading and Differences in Text Reading Fluency in Adult Readers. PLoS ONE 8:8.

Wang XF, Chen G. 2003. Complex Networks: small-world, scale-free and beyond. Circuits and Systems Magazine, IEEE 3:6–20.

Webber CL, Marwan N, Facchini A, Giuliani A. 2009. Simpler methods do it better: Success of Recurrence Quantification Analysis as a general purpose data analysis tool. Physics Letters A 373:3753–3756.

Webber Jr. CL, Zbilut JPP, Webber Jr CL. 2005. Recurrence quantification analysis of nonlinear dynamical systems. Tutorials in contemporary nonlinear methods for the behavioral sciences:26–94.

Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 7(1), 49–63. doi:10.1016/S0163-6383(84)80022-3

Werker, J., & Tees, R. (1987). Speech perception in severely disabled and average reading children. Canadian Journal of Psychology/Revue, (41), 48–61.

Whiteley, H., & Smith, C. (2001). The use of tinted lenses to alleviate reading difficulties. Journal of Research in Reading, 24(1), 30–40. doi:10.1111/1467-9817.00131

Wijnants, M. L., Bosman, A. M. T., Hasselman, F., Cox, R. F. A., & Van Orden, G. C. (2009). 1/f scaling in movement time changes with practice in precision aiming. Nonlinear Dynamics, Psychology, and Life Sciences, 13(1), 79–98.

Wijnants, M. L., Cox, R. F. A., Hasselman, F., Bosman, A. M. T., & Van Orden, G. C. (2012). A trade-off study revealing nested timescales of constraint. Frontiers in Physiology, 3(May), 116. doi:10.3389/fphys.2012.00116

Wijnants, M. L., Hasselman, F., Cox, R. F. A., Bosman, A. M. T., & Van Orden, G. (2012). An interaction-dominant perspective on reading fluency and dyslexia. Annals of dyslexia, 62(2), 100–19. doi:10.1007/s11881-012-0067-3

Wilkinson, L. (1999). Statistical Methods in Psychology Journals Guidelines and Explanations. American Psychologist, 54(1988), 1–18.

Wimmer, H., & Schurz, M. (2010). Dyslexia in regular orthographies: manifestation and causation. Dyslexia, 299, 283–299. doi:10.1002/dys

Worrall, J. (1989). Structural realism: The Best of Both Worlds?. dialectica, 43(1-2), 99–124. doi:10.1111/j.1746-8361.1989.tb00933.x

Worrall, J. (2011). Causality in medicine: getting back to the Hill top. Preventive Medicine, 53(4-5), 235–8. doi:10.1016/j.ypmed.2011.08.009

Yasnitsky, A. (2011). Lev Vygotsky: Philologist and defectologist, A Socio-Intellectual Biography. In Portraits of pioneers in developmental psychology (pp. 105–129).

Yates, F. (1968). Theory and Practice in Statistics. Journal of the Royal Statistical Society, 131(4), 463–477.

Yilmaz E, Uzuntarla M, Ozer M, Perc M. 2013. Stochastic resonance in hybrid scale-free neuronal networks. Physica A: Statistical Mechanics and its Applications 392:5735–5741.

Zavershneva, e. I. (2012). Investigating L. S. Vygotsky's Manuscript "The Historical Meaning of the Crisis in Psychology." Journal of russian and east european Psychology, 50(4), 42–63. doi:10.2753/rPO1061-0405500402

Zbilut JP, Giuliani A, Webber CL. 1998. Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. Physics Letters A 246:122–128.

Ziegler JC, Goswami U. 2005. Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. Psychological bulletin 131:3–29.